Facultad de Psicología
Departamento de Psicología Social y Metodología

Dimensionality Assessment of Ordinal Variables: An Evaluation of Classic and Modern Methods

---

Doctoral Dissertation

Author:

Luis Eduardo Garrido de los Santos

Directors:

Vicente Ponsoda Gil

Francisco José Abad

Madrid, June 2012

A mis padres, Cristina y Carlos

Gracias por haberme dado esta gran oportunidad y por todo el amor y apoyo que me han brindado en estos años. ¡No lo podría haber hecho sin ustedes! Los quiero mucho

ABSTRACT

The aims of this study are to first evaluate the performance of factor retention criteria in the dimensionality assessment of ordinal variables, and, second, to offer clear and easy-to-follow guidelines for researchers who work with ordinal-level data in practice. The determination of the number of factors is considered to be a crucial decision within the context of exploratory factor analysis (EFA) and structural equation modeling (SEM), but unfortunately, has been largely overlooked as it pertains to ordinal observed variables, which are typically encountered in the social and behavioral sciences. The current study seeks to address this issue by taking an in-depth look at the performance of three "classic" factor retention criteria (Velicer's Minimum Average Partial method [MAP], Horn's Parallel Analysis [PA], and the eigenvalue-greater-than-1 rule [K1]), as well as four fit indices (the Comparative Fit Index [CFI], Tucker-Lewis Index [TLI], Root Mean Square Error of Approximation [RMSEA], and the Standardized Root Mean Square Residual [SRMR]), in the dimensionality assessment of ordinal variables.

In order to broadly evaluate the accuracy of the factor retention criteria, a comprehensive set of factors was systematically manipulated using Monte Carlo methods, including the factor loading, number of variables per factor, number of factors, factor correlation, sample size, number of response categories, level of skewness, extraction method, and type of correlation matrix. The results showed that PA with principal component analysis, polychoric correlations, and the mean eigenvalue criteria, along with the CFI and TLI indices at a cutoff value of 0.95, perform adequately in determining the number of factors with ordinal variables. The other four methods, however, could not be recommended due to the strong levels of bias they exhibited. The performances of the factor retention criteria are put into theoretical context and guidelines are offered on how to assess the dimensionality of ordinal-level data in practice.

RESUMEN

Los objetivos de este estudio son, primero, evaluar el rendimiento de los criterios de retención de factores en la estimación de dimensionalidad para variables ordinales, y, segundo, ofrecer a los investigadores guías prácticas que sean claras y fáciles de seguir. La determinación del número de factores es considerada como una decisión crucial en el análisis factorial exploratorio (EFA) y en los modelos de ecuaciones estructurales (SEM), mas sin embargo, se ha pasado por alto en gran medida en lo que concierne a variables ordinales, encontradas comúnmente en las ciencias sociales y del comportamiento. Mediante el estudio actual se busca abordar este problema a partir de una evaluación extensiva de tres métodos "clásicos" (el método de las Mínimas Correlaciones Parciales Medias [MAP], el Análisis Paralelo [PA] y la regla de los autovalores mayores a 1 [K1]), así como de cuatro índices de ajuste (el Índice de Ajuste Comparativo [CFI], el Índice de Tucker-Lewis [TLI], la Raíz Media Cuadrática del Error de Aproximación [RMSEA] y la Raíz Media Cuadrática del Residual Estandarizado [SRMR]), en la estimación de dimensionalidad para datos ordinales.

Con el fin de evaluar de una manera amplia la efectividad de las reglas de retención de factores, se manipuló un grupo integral de factores a partir de métodos Monte Carlo, incluyendo las saturaciones, el número de variables por factor, número de factores, correlaciones factoriales, tamaño muestral, número de categorías de respuesta, nivel de asimetría, método de extracción y el tipo de matriz de correlación. Los resultados mostraron que PA con componentes principales, correlaciones policóricas y la media de los autovalores aleatorios, en conjunto con los índices CFI y TLI con un punto de corte de 0.95, tienen un desempeño adecuado en la determinación del número de factores para variables ordinales. Los otros cuatro métodos, sin embargo, no pudieron recomendarse debido a los altos niveles de sesgo que mostraron. El desempeño de los criterios de retención es puesto en un contexto teórico y se ofrecen guías prácticas para determinar la dimensionalidad de datos ordinales.

TABLE OF CONTENTS

1. ACKNOWLEDGMENTS

Quisiera agradecer de manera especial a mis directores de tesis, Vicente Ponsoda y Paco Abad, por haberme proporcionado una experiencia formativa del más alto nivel. Me siento privilegiado por haber sido vuestro estudiante. A través de estos años he aprendido muchas cosas de ustedes, tanto a nivel profesional como personal, y aprecio mucho la disposición, el entusiasmo, la paciencia, el apoyo, y las incontables risas que me brindaron desde que les conocí. ¡Muchísimas gracias!

Muchas gracias también a…

Julio por haber apostado por mí.

Agustín por una amistad de las que hay muy pocas en la vida y por las incontables cosas con las que me ayudaste en mi estadía en Madrid.

A Fabiola por tantas experiencias compartidas durante todos esos años… ¡Cuantos recuerdos!

A Esther, Jose y José Luis por ser mi familia en la residencia.

A mis hermanos Carlos Emilio, Raquel y Guillermo, por el cariño y apoyo que me cada uno y que me llena el corazón.

2. GENERAL INTRODUCTION

Researchers in the social sciences are often interested in the study of latent – unobserved – variables that are the causes of the behaviors they observe (Bollen, 2002). Latent variables such as aptitudes, feelings, and motives, in the context of a well-reasoned theory, have the potential to explain a wide array of behavioral processes using a relatively small number of constructs (Hoyle & Duvall, 2004). These latent variables, or factors, are random variables whose properties must be inferred indirectly using a statistical model that connects the latent variables to manifest – observed – variables, believed to be caused, at least in part, by one or more factors (Mulaik & Millsap, 2000). The primary statistical tool for drawing such inferences is factor analysis, a technique that aims to describe the associations among a potentially large number of observed variables using a relatively small number of factors, the latent variables in the statistical model that represent the underlying processes that have been operating (Browne & Cudeck, 1992; Hoyle & Duvall, 2004).

A fundamental concern in exploratory factor analysis (EFA), and in extension, structural equation modeling (SEM), is the determination of the number of factors to retain for a group of variables of interest (Fabrigar, Wegener, MacCallum, & Strahan, 1999; Hayduk & Glaser, 2000; Hayton, Allen, & Scarpello, 2004; Henson & Roberts, 2006; Mulaik & Millsap, 2000; Schmitt, 2011). Extracting too few factors (i.e., underfactoring) will likely result in substantial error, as multiple factors may be combined and variables may load on the wrong factors (Fava & Velicer, 1996; Wood, Tataryn, & Gorsuch, 1996). Alternatively, extracting too many factors (i.e., overfactoring) can result in "factor splitting" as well as factors that are non-interpretable or unreliable (Fava & Velicer, 1992; Lee & Comrey, 1979; Wood et al., 1996). Although overfactoring appears to have less deleterious effects than underfactoring (Fabrigar et al., 1999), both types of misspecifications will lead to poor factor loading pattern reproduction and interpretation (Velicer, Eaton, & Fava, 2000), and can

potentially misleading theory development efforts (Fabrigar et al., 1999; Patil, Singh, Mishra, & Donovan, 2008). Moreover, researchers have recently embedded the dimensionality decision within the broader SEM framework, as an erroneous specification of the number of factors can lead to the rejection of otherwise adequately conceived SEM models (Mulaik & Millsap, 2000). Nevertheless, despite the substantial attention that this topic has received for more than half a century (e.g., Cattell, 1966; Horn, 1965; Kaiser, 1960; Steiger & Lind, 1980; Timmerman & Lorenzo-Seva, 2011; Velicer, 1976; Velicer,et al., 2000), it continues to be a subject of considerable debate among quantitative methodologists (e.g., Bollen, 2000; Hayduk & Glaser, 2000; Herting & Costner, 2000; Mulaik & Millsap, 2000) and remains poorly understood by applied researchers (Fabrigar et al., 1999; Hayton et al., 2004; Henson & Roberts, 2006).

There are many rules for deciding the number of factors to retain, and they can be divided into three categories (Floyd & Widaman, 1995): statistical tests, mathematical and psychometric criteria, and rules of thumb.

Statistical tests. Statistical tests for EFA and SEM are available for some estimation methods such as maximum likelihood, generalized least squares, and asymptotically distribution free estimation, and are computed as chi-square (χ^2) significance tests of the residual covariation among observed variables after extracting a certain number of factors. If the chi-square statistic is significant, the model is rejected in favor of a model with one or more additional factors (Floyd & Widaman, 1995; Fabrigar et al., 1999).

Mathematical and psychometric criteria. These factor retention methods are more commonly used and include the Principal Component Analysis' (PCA) based eigenvalue-greater-than-1 rule or Kaiser-Guttman criterion (**K1**; Kaiser, 1960), Parallel Analysis (**PA**; Horn, 1965), and the Minimum Average Partial method (**MAP**; Velicer, 1976), as well as the

multiple fit indices that have been proposed in the context of SEM modeling (e.g., Bentler, 1990, 1995; Bentler & Bonnet, 1980; Steiger & Lind, 1980). The K1 rule is the default in most statistical packages and is based on population proofs and psychometric arguments regarding the size of the eigenvalues for uncorrelated variables. Parallel Analysis, for its part, can be considered as a "sample version" of the K1 rule because it is based on the same population proofs and psychometric arguments, but unlike the K1 rule, it also takes into account sampling error, and least squares "capitalization" on this error in the computation of the latent roots (Horn, 1965). In addition, the MAP method employs a matrix of partial correlations, separating the common and unique variance and retaining those factors that consist primarily of common variance (Velicer, 1976; Velicer et al., 2000). On the other hand, fit indices, which are usually based on the χ^2 statistic and measure the "practical usefulness" of statistical models (Bentler & Bonnet, 1980), have also been recommended in recent years for the assessment of data dimensionality (Fabrigar et al., 1999).

Rules of thumb. Many practical criteria fall under the rubric of rules of thumb, like the scree test (Cattell, 1966), the percent of total variance accounted for, and the ratio of first-to-second eigenvalues. Of these factor retention rules, the scree test is by far the most commonly used and studied (Fabrigar et al., 1999; Hayton et al., 2004), and it is based on the visual inspection of the plot of the eigenvalues corresponding to the unrotated factors.

2.1. Description of the Most Relevant Factor Retention Methods

The most relevant factor retention methods will be described in this section. These particular methods have been chosen because they have performed well in previous Monte Carlo dimensionality studies (e.g., Cho, Li, & Bandalos, 2009; Timmerman & Lorenzo-Seva, 2011; Zwick & Velicer, 1982, 1986; Velicer et al., 2000) and/or because they are often used in practice (Fabrigar et al., 1999; Hayton et al., 2004). In order to give a historical context,

the factor retention methods will be presented according to the chronological order by which they were originally proposed.

2.1.1. The eigenvalue-greater-than-1 rule (1960)

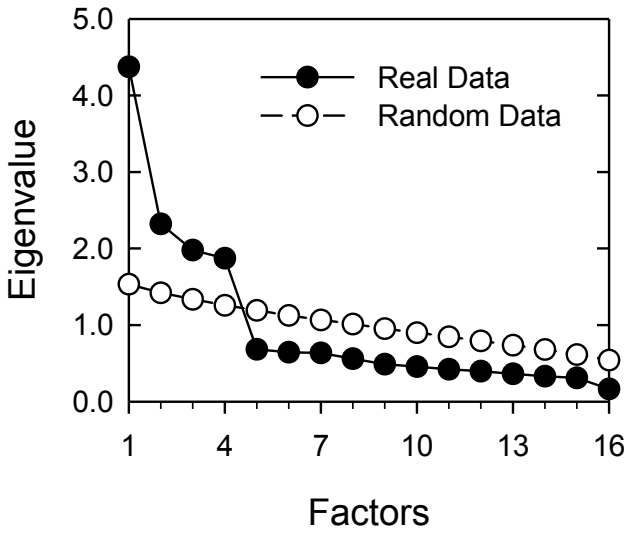
The K1 rule posits that only factors with eigenvalues greater than 1 should be retained. This criterion is based on Kaiser's (1960) proofs and arguments that Guttman's (1954) latent-root-one lower bound estimate for the minimum rank of a correlation matrix could be used as a psychometric upper bound for the number of factors problem. The rationale behind this rule is that a factor should be able to explain at least as much variance as a variable is accorded in the standard score space (Dickman, 1960), and that a threshold of 1 ensures that the component will have a positive internal consistency (Kaiser, 1960).

2.1.2. Parallel Analysis (1965)

Horn (1965) proposed the PA method on the basis of Kaiser's (1960) and Dickman's (1960) proofs and arguments that Guttman's (1954) latent-root-one lower bound estimate for the minimum rank of a correlation matrix could be used as a psychometric upper bound for the number of factors problem. Because the proofs for the eigenvalue-greater-than-1 rule were performed on population statistics, Horn (1965) argued that due to sampling error, and least squares "capitalization" on this error in the computation of the latent roots, some components from uncorrelated variables in the population could have eigenvalues > 1 at the sample level. Therefore, he proposed PA as a means to estimate and take into account the proportion of variance that was due to sampling error and chance capitalization. In this sense, PA may be viewed as a sample alternative to the eigenvalue-greater-than-1 rule. Instead of retaining factors that have eigenvalues > 1, with PA only those factors that have eigenvalues greater than those generated from independent variates are retained. The goal is to account

for chance capitalization in the sample eigenvalues under the null hypothesis of independent variables (Buja & Eyuboglu, 1992).

The implementation of the PA procedure involves the generation of a large number of matrices of random data. Each matrix is generated with the same number of subjects and variables as the real data matrix under assessment. Then, the number of factors is determined by comparing the eigenvalues from the real data matrix with the mean of the eigenvalues from the random data matrices (Horn, 1965). A factor is retained as long as its eigenvalue is greater than the mean eigenvalue from its random counterpart. An example of the PA method is presented in Figure 2.1. In this case, a 4-factor solution is suggested.



Note. The values for the random data are the mean eigenvalues across 100 samples

Figure 2.1: *An illustration of the Parallel Analysis Method*

2.1.3. The Scree Test (1966)

The scree test (Cattell, 1966) plots the eigenvalues of the unrotated factors on a coordinate plane and examines the slope of the line connecting them (Floyd & Widaman, 1995). The rationale for this test is that a few major factors account for the most variance,

resulting in an initial steep "cliff" in the plot of the eigenvalues, followed by a shallow "scree" describing the relative constant and non-systematic variance that is accounted for by the numerous minor factors (Hayton et al., 2004). The number of factors to retain is equal to the number of factors that do not belong to the scree. An example of the scree test is shown in Figure 2.2. In this case, the scree test suggests a 4-factor structure, as only the first 4 factors are located before the "break" in the curve and thus do not belong the scree or flat portion of the curve.

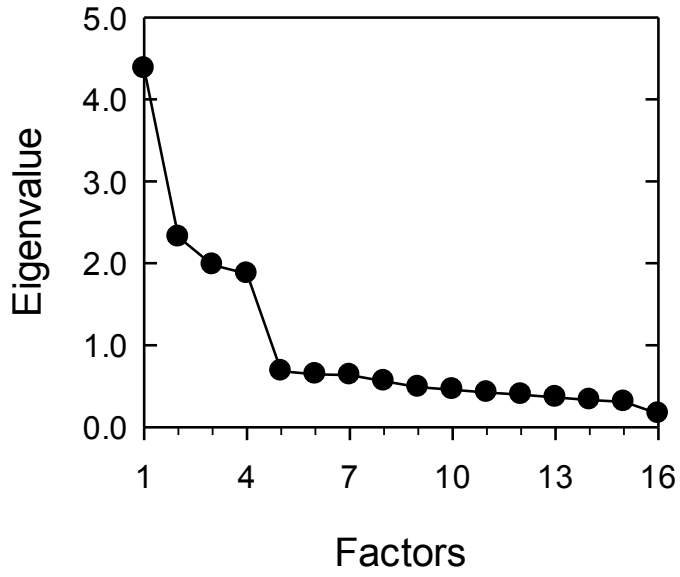


Figure 2.2: *An illustration of the Scree Test*

2.1.4. The Minimum Average Partial Method (1976)

The MAP method (Velicer, 1976) was developed in the context of PCA, and is based on the matrix of partial correlations. Each component is partialled out of the correlation matrix and the average of the squared partial correlations is computed. The number of factors to retain is determined by the point where the minimum average of the squared partial correlations is obtained. The rationale of this procedure can be described as follows: as common variance is partialled out of the correlation matrix for each successive component,

the MAP criterion will keep on decreasing. At the point where the common variance has been removed, extracting additional components will result in unique variance being partialled out, and the MAP criterion will begin to rise. The MAP procedure, therefore, provides an unequivocal stopping point for the number of factors by separating the common and unique variance and retaining only those factors that consist primarily of common variance. An illustration of the MAP method is presented in Figure 2.3. In this case, the MAP method suggests a 2-factor structure as the minimum MAP criterion is obtained after partialling out the first two components.

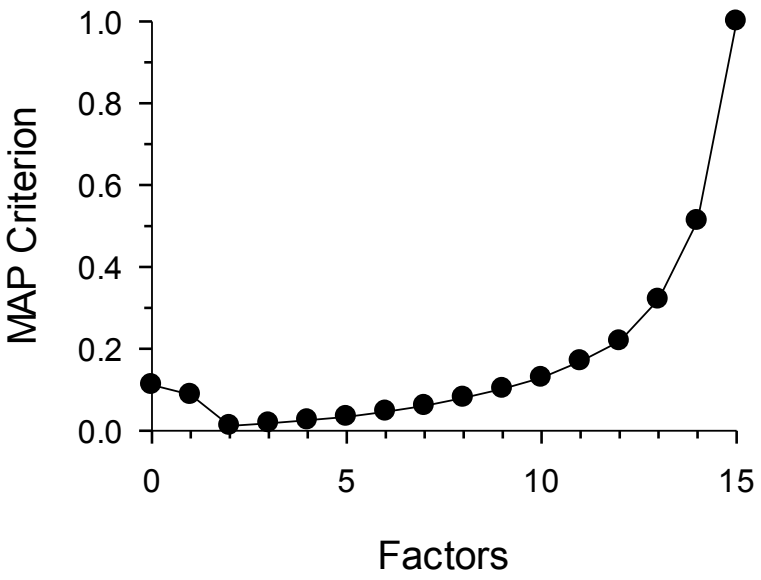


Figure 2.3: *An illustration of the MAP method*

The MAP procedure begins with the computation of the partial covariance matrix,

$$\mathrm{C}_m = \mathrm{R} - \mathrm{A}_m\mathrm{A}_m^{\mathrm{T}} \tag{2.1}$$

where C_m is the partial covariance matrix that results from partialling out the first m components from R, R is the correlation matrix, and A_m is the component loading matrix for components 1 to m.

Next, the partial correlation matrix is obtained

$$\mathrm{R}_m^* = \mathrm{D}^{-\frac{1}{2}}\,\mathrm{C}_m\,\mathrm{D}^{-\frac{1}{2}} \tag{2.2}$$

where R_m^* is the partial correlation matrix, and

$$\mathrm{D} = \mathrm{diag}(\mathrm{C}_m) \tag{2.3}$$

The MAP criterion is then obtained by averaging the squares of the partial correlations contained in R_m^*

$$\mathrm{MAP}_m = \sum_{\mathrm{i}=1}^{p}\sum_{\substack{\mathrm{j}=1 \\ \mathrm{i}\neq\mathrm{j}}}^{p} \frac{{\mathrm{r}_{\mathrm{ij}}}_m^{*\,2}}{p(p-1)} \tag{2.4}$$

where p is the number of variables.

This procedure is repeated until $p-1$ components have been partialled out of R (partialling out p components would result in a null partial covariance matrix). Finally, Velicer proposes a test for the first factor by averaging the squares of the correlations contained in R

$$\mathrm{MAP}_0 = \sum_{\mathrm{i}=1}^{p}\sum_{\substack{\mathrm{j}=1 \\ \mathrm{i}\neq\mathrm{j}}}^{p} \frac{\mathrm{r}_{\mathrm{ij}}{}^2}{p(p-1)} \tag{3.5}$$

if $\mathrm{MAP}_0 < \mathrm{MAP}_1$, no factors should be extracted.

2.1.5. Fit Indices (1980-)

The general approach for estimating data dimensionality with fit indices is to fit a 1-factor model to the correlation matrix and if the index or indices in question show a lack of it, more factors are sequentially extracted until an acceptable level of fit is reached (Asparouhov

& Muthén, 2009). The number of factors is determined at the point where an acceptable fit is first obtained. A brief summary of the rationale behind the use fit indices to evaluate model fit is presented below.

Large-sample theory provides a chi-square (χ^2) goodness-of-fit test for comparing a model against a general alternative model based on correlated variables (Bentler & Bonett, 1980). The asymptotic χ^2 statistic tests the null hypothesis that the population covariance matrix of observed variables is equal to the model-implied covariance matrix (Hu & Bentler, 1999). However, because in the social sciences any model is at best an approximation to reality, the null hypothesis of exact fit is known *a priori* to be false, resulting invariably, with a large enough sample, in the rejection of even models that closely approximate the population covariance matrix (Bentler & Bonett, 1980; Browne & Cudeck, 1992; Schermelleh-Engel, Moosbrugger, & Müller, 2003; Yu, 2002). In order to overcome these limitations of the χ^2 statistical test of exact fit, Bentler and Bonett (1980) introduced the concepts of fit indices and practical significance in relation to the assessment of latent variable models. Bentler and Bonett (1980) proposed the use of *incremental* fit indices as a means to compute the amount of information gained when comparing competing models. They argued that an index of information gained would provide valuable information about the practical usefulness of competing models and should be independent of sample size and statistical significance test information (Bentler & Bonett, 1980). Incremental fit indices assess the degree to which the tested model is superior to an alternative "baseline" model in reproducing the observed covariance matrix. The baseline model is usually a null model in which all the observed variables are uncorrelated (Hu & Bentler, 1999).

In addition to the incremental fit indices, measures of absolute fit have also been developed to evaluate the appropriateness of latent variable models. The *absolute* fit indices

measure the degree to which the hypothesized model corresponds to the empirical data. In this case, no reference model is used to assess the amount of increment in model fit, but an implicit comparison is made to a saturated model that exactly reproduces the covariance matrix of observed variables (Hu & Bentler, 1999). Some of the most commonly used fit indices are the Comparative Fit Index (**CFI**; Bentler, 1990), Tucker-Lewis Index or Non-Normed Fit Index (**TLI**; Tucker & Lewis, 1973), Root Mean Square Error of Approximation (**RMSEA**; Steiger & Lind, 1980), and the Standardized Root Mean Square Residual (**SRMR**; Bentler, 1995).

2.2. Review of the Factor Retention Literature

2.2.1. Dimensionality Assessment of Continuous Variables

The accuracy of factor retention criteria in the dimensionality assessment of continuous variables has focused primarily on the following 4 "classic" methods: the K1 rule, the scree test, PA, and the MAP method. A brief summary of the most relevant findings from this literature is presented next.

Zwick and Velicer (1982) used Monte Carlo methods to evaluate the effectiveness of the K1 rule, Bartlett's significance test (Bartlett, 1950, 1951), the scree test, and the MAP method. The results of their study showed that the scree test was the most accurate method, followed by the MAP procedure, which had a tendency to underfactor with a small number of variables per factor. Next in line was Bartlett's test, which was found to be adequate except with large numbers of variables, and then the K1 rule, which had a tendency to severely overestimate the number of factors. In general, the factor loading was the independent variable that had the greatest impact in the performance of the factor retention methods evaluated in this study.

Using more complex data, Zwick and Velicer (1986) found PA to be more accurate than the MAP method, which in turn, was superior to the scree test, Bartlett's test, and the K1 rule. In this study Bartlett's test was found to be very variable and particularly sensitive to the sample size, the number of variables, and the level of saturation. In general, this method could not be recommended due to its strong tendency to retain trivial factors. Also, and in line with previous studies, the K1 rule was found to grossly overestimate the number of factors, especially with large numbers of observed variables. The scree test, for its part, tended to overfactor with low levels of saturation, but on the whole, performed adequately. Based on these results, the scree test was recommended as an adjunct to the PA and MAP methods. Regarding the MAP method, the same tendency to underfactor with low factor loadings and a small number of variables per factor was found in this study, replicating previous results. On the other hand, PA was consistently the most accurate method, and its performance improved with increases in sample size, level of saturation, and number of variables per factor.

Velicer et al. (2000) conducted a simulation study that evaluated the K1 rule, PA, and the MAP method. In this study, the authors introduced a variant to the MAP procedure by raising the partial correlations to the 4^{th} power instead of squaring them. The results from their simulation study showed, again, that PA was the most accurate method, followed closely by the MAP method. Furthermore, the new variant of MAP with the 4^{th} power produced slightly more accurate estimations than the original MAP method. Also, as in previous studies, the K1 rule tended to greatly overfactor and was the most variable and least accurate factor retention method. Finally, a simulation study by Peres-Neto, Jackson, and Somers (2005) evaluated the performance of 20 stopping rules, and based on the results from this study they recommended, among others, the PA and MAP methods. In line with previous findings, the MAP method showed a strong tendency to underfactor when it was in error.

Based mostly on these simulation studies with continuous variables, numerous authors have recommended the use of the PA and MAP methods to assess data dimensionality (Fabrigar et al., 1999; Floyd & Widaman, 1995; Hayton et al., 2004; Henson & Roberts, 2006; Patil et al., 2008). However, even though these recommendations are based on the findings with continuous variables, researchers have often used these methods without modification for the determination of the number of factors with ordinal variables (e.g., Eklöf, 2006; Lai, Crane, & Cella, 2006; Wood, Maltby, & Joseph, 2008), where their applicability is in question and their efficacy has not been established.

2.2.2. Dimensionality Assessment of Ordinal Variables

Variables characterized by an ordinal level of measurement are common in many empirical investigations within the social and behavioral sciences (Flora & Curran, 2004). One of the most typical situations that generate ordinal variables is the use of psychometric tests that contain Likert scales (e.g., strongly disagree to strongly agree). Although the individual items are designed to measure continuous phenomena, the observed responses are discrete realizations of a small number of categories (Flora & Curran, 2004; Olsson, 1979a). If these observed ordinal variables are analyzed with statistical methods that assume continuous distributions, there is the potential for critical mismatch between the assumptions underlying the statistical model and the characteristics of the data to be analyzed (Bollen & Barb, 1981; Flora & Curran, 2004). Because of this, it is usually recommended that ordinal variables be analyzed with methods specifically developed for this type of measurement level (Savalei & Rhemtulla, in press).

Several factors make the dimensionality assessment of ordinal data more difficult than for normally distributed continuous variables. As it is well known, Pearson's product-moment correlation underestimates the strength of the relationship between ordinal variables

(Babakus, Ferguson, & Jöreskog, 1987; Bollen & Barb, 1981), and may produce spurious dimensions known as "difficulty factors" when the variables are skewed in opposite directions (Gorsuch, 1983; Olsson, 1979b). Because of these biases, the polychoric correlation coefficient has been recommended as a measure of association for the factor analysis of ordinal variables (Flora & Curran, 2004; Jöreskog & Moustaki, 2001). Assuming that the ordinal variables are a crude measure of underlying bivariate normally distributed variables, the polychoric correlation is a maximum likelihood estimate of the Pearson correlation between the underlying variables (Olsson, 1979a). Polychoric correlations have been shown to produce unbiased parameter estimates for both exploratory and confirmatory factor analysis (Babakus et al., 1987; Flora & Curran, 2004). Despite these advantages, however, polychoric correlations have some problems of their own. In particular, they frequently produce non-Gramian correlation matrices (matrices with negative eigenvalues), have large sampling errors, and can take considerable time to estimate, properties that can potentially compromise the effectiveness and applicability of factor retention methods such as PA or MAP (Timmerman & Lorenzo-Seva, 2011; Tran & Formann, 2009; Weng & Cheng, 2005).

The research on the dimensionality assessment of ordinal variables has focus primarily on the PA method. A brief overview of this factor retention literature is presented next. Weng and Cheng (2005) first studied the effectiveness of PA with unidimensional binary data by comparing empirical eigenvalues from Pearson and tetrachoric correlations with random eigenvalues from a multivariate normal distribution. The results from their simulation study with positively skewed variables showed that PA with Pearson correlations (PAr) was more accurate than PA with polychorics (PAρ), a finding they attributed to the large sampling errors and unstable behavior of the tetrachoric correlations. A subsequent study by Tran and Formann (2009) extended the evaluation of PA with unidimensional binary data by

simulating factors with both positively and negatively skewed items, the scenario most likely to produce difficulty factors with Pearson correlations. The results from their study indicated that neither PAr nor Paρ could be recommended; in the case of PAr, because of poor performance, and in the case of Paρ, due to applicability issues stemming from the large number of non-Gramian polychoric matrices (indefinite matrices that have at least one negative eigenvalue).

Cho et al. (2009) further advanced the study of PA with ordinal variables by assessing its performance with polytomous items of 2 and 3 response options and multidimensional structures of symmetrically distributed variables. In this case, the authors matched the type of correlation matrix used to compute the real and random eigenvalues and found that PAr was at least as accurate as Paρ. More recently, Timmerman & Lorenzo-Seva (2011) proposed a new version of PA with Minimum Rank Factor Analysis (MRFA), and studied its effectiveness, along with the original version of PA, with multidimensional structures of polytomous items skewed in opposite directions. According to their results, PA with MRFA extraction was moderately superior to PA with PCA extraction. Also, these authors noted that Paρ could only be computed for 37% of the data matrices due to convergence problems of the algorithm used to compute the polychoric correlations or because the smoothing procedure did not produce a Gramian polychoric matrix. This situation prompted them to state that the "convergence problems of the polychoric approach prevent its general application to empirical data" and "may pose severe problems" in practice (p. 218).

As can be seen from the preceding commentary, the performance of PA with ordinal variables has been difficult to ascertain due to the inconclusive and often contradictory findings from previous studies. For example, some studies have suggested that PA works better with Pearson correlations (Weng & Cheng, 2005), others that PA performs equally well with Pearson and polychoric correlations (Cho, Li, & Bandalos, 2009), and still others

that PA is ineffective with either type of correlation matrix (Tran & Formann, 2009). Also, several studies suggest that PA has severe applicability issues with the more theoretically appropriate polychoric correlations (Timmerman & Lorenzo-Seva, 2011; Tran & Formann, 2009). Furthermore, a new version of PA with MRFA has been recently proposed (Timmerman & Lorenzo-Seva, 2011), and preliminary analyses have shown that it may be moderately superior to the original PCA based PA. In addition to the inconclusive and contradictory findings from the literature with the PA method, other frequently recommended criteria such as the MAP method (Henson & Roberts, 2006; Velicer et al., 2000) have not been evaluated in the context of ordinal-level data.

Fit indices constitute another set of factor retention criteria that may be employed with ordinal variables. Using fit indices for dimensionality assessment may be advantageous because it would give researchers access to important model diagnostic information, such as the presence of correlated errors (Bollen, 2000), when making the determination of the number of factors. Moreover, it might reduce the need for ad-hoc model manipulation in the more advanced stages of testing, such as the evaluation of a full-blown SEM model, due to an erroneous specification of the number of factors (Mulaik & Millsap, 2000). However, despite the fact that fit indices have been recommended and used with increased frequency in recent years for dimensionality assessment (e.g., Asparouhov & Muthén, 2009; Browne & Cudeck, 1992; Hoyle & Duvall, 2004; Fabrigar et al., 1999; Floyd & Widaman, 1995; Steiger & Lind, 1980; Tepper & Hoyle, 1996), their efficacy with ordinal variables has yet to be systematically evaluated.

2.3. Goals of the Current Dissertation

The main goal of this dissertation is to evaluate the performance of "classic" and "modern" factor retention criteria in the determination of the number of factors with ordinal

variables across a comprehensive set of factors. The classic methods that will be evaluated include the K1 rule, PA, and the MAP method. On the other hand, the modern factor retention criteria are composed of two incremental fit indices, CFI and TLI, as well as two absolute fit indices, RMSEA and SRMR. A secondary goal of this dissertation is to offer clear and easy-to-follow practical guidelines to researchers who work with ordinal data.

Because many of these factor retention methods require special modifications and present unique challenges when they are used for the dimensionality assessment of ordinal-level data, three separate Monte Carlo studies will be carried out in order to evaluate each of the methods thoroughly. The specific goals of the three Monte Carlo studies are presented below.

2.3.1. Study 1: The MAP Method

The main goal of Study 1 is to determine which variant of the MAP method performs best with ordinal variables. In this line, two "method" factors will be manipulated: the type of correlation matrix (Pearson or polychoric) and the power the partial correlations are raised to (2^{nd} or 4^{th}). This study will also determine the saliency of multiple independent variables such as the factor loading and the number of variables per factor in the accuracy of the MAP procedure. In addition, practical guidelines will be offered on how to use the MAP method with ordinal variables. An issue that needs to be addressed in this study is how to deal with the non-Gramian polychoric matrices, as the MAP method cannot be computed with these types of matrices.

2.3.2. Study 2: Parallel Analysis

The main goal of Study 2 is to determine which variant of the PA method performs best with ordinal variables. In this case, three method factors will be manipulated: the type of

correlation matrix (Pearson or polychoric), the extraction method (PCA or MRFA), and the percentile of the random eigenvalues (the mean or the 95th percentile). As part of the evaluation of the PA method, an attempt will be made to identify and resolve the issues that have produced the inconsistent results found in the literature. In addition, this study will also determine the saliency of multiple independent variables in the accuracy of the PA method and will offer practical guidelines on how to use PA with ordinal variables. An issue that also needs to be addressed with PA is how to deal with the non-Gramian polychoric matrices, which most authors have discarded in the past.

2.3.3. Study 3: Fit Indices

The main goal of Study 3 is to evaluate the performance of various fit indices in the determination of the number of factors with ordinal variables. In order to give context to the performance of the fit indices, three classic factor retention methods will also be included in the simulation: the K1 rule, PA, and the MAP method. In the case the PA and MAP methods, only the most accurate variants, established through Studies 1 and 2, will be evaluated. A secondary goal is to establish the optimal cutoff values for each fit index. In addition, this study will determine the saliency of multiple independent variables in the accuracy of the fit indices and will offer practical guidelines for the dimensionality assessment of ordinal variables based on classic and modern factor retention criteria. An issue that needs to be addressed in this study is how to deal with non-converged factor solutions.

3. STUDY 1: PERFORMANCE OF VELICER'S MAP FACTOR RETENTION METHOD WITH ORDINAL VARIABLES

Despite strong evidence supporting the use of Velicer's minimum average partial (MAP) method to establish the dimensionality of continuous variables, little is known about its performance with ordinal data. Seeking to fill this void, the current study takes an in-depth look at the performance of the MAP procedure in the presence of ordinal-level measurement. Using Monte Carlo methods, seven factors related to the data were systematically manipulated (sample size, factor loading, number of variables per factor, number of factors, factor correlation, number of response categories, and skewness) as well as two factors related to the MAP method (type of correlation matrix and power). The results indicate that using polychoric correlations and the squared partial correlations leads to considerably more accurate estimations than using Pearson correlations and/or raising the partial correlations to the 4^{th} power. Additionally, the MAP method is shown to be a biased estimator of dimensionality in two conditions: (1) for low factor loadings (0.40); and (2) for medium factor loadings (0.55) and a small number of variables per factor (≤ 6). The applicability of this method with ordinal variables is discussed in the context of these findings.

Determining the number of factors to retain is one of the most critical phases of an exploratory factor analysis (Fabrigar, Wegener, MacCallum, & Strahan, 1999; Hayton, Allen, & Scarpello, 2004; Henson & Roberts, 2006; Velicer, Eaton, & Fava, 2000). Extracting too few factors (i.e., underfactoring) will likely result in substantial error, as multiple factors may be combined and variables may load on the wrong factors (Fava & Velicer, 1996; Wood, Tataryn, & Gorsuch, 1996). Alternatively, extracting too many factors (i.e., overfactoring) can result in "factor splitting" as well as factors that are non-interpretable or unreliable (Fava

& Velicer, 1992; Lee & Comrey, 1979; Wood et al., 1996). Although overfactoring appears to have less deleterious effects than underfactoring (Fabrigar et al., 1999), both types of miss-specifications will lead to poor factor loading pattern reproduction and interpretation (Velicer at al., 2000), and can potentially mislead theory development efforts (Patil, Singh, Mishra, & Donovan, 2008).

Numerous factor retention methods have been proposed throughout the long history of factor analysis (see Peres-Neto, Jackson, & Somers, 2005; Velicer et al., 2000), including the eigenvalue-greater-than-1 rule (Kaiser, 1960), the scree test (Cattell, 1966), parallel analysis (Horn, 1965), and the minimum average partial method (MAP; Velicer, 1976). The results from simulation studies with continuous variables have shown that parallel analysis and MAP are the two most accurate methods (Peres-Neto et al., 2005; Velicer et al., 2000; Zwick & Velicer, 1986), followed by the scree test and the eigenvalue-greater-than-1 rule, which is not currently recommended due to its poor performance and flawed rationale (Cliff, 1988; Zwick & Velicer, 1982, 1986). In contrast, factor retention studies with ordinal variables have so far been scarce and inconclusive (see Cho, Li, & Bandalos, 2009; Tran & Formann, 2009; Weng & Cheng, 2005).

The lack of factor retention research with ordinal variables is problematic because the findings for continuous data may not apply to analyses with ordinal-level measurement. For instance, it is well known that Pearson's product-moment correlation, used for continuous data, underestimates the strength of the relationship between ordinal variables (Babakus, Ferguson, & Jöreskog, 1987; Bollen & Barb, 1981), and may produce spurious dimensions known as "difficulty factors" when the variables are skewed in opposite directions (Gorsuch, 1983; Olsson, 1979b). Because of this, Pearson correlations are expected to produce biased dimensionality estimates for ordinal data. The polychoric correlation coefficient, on the other hand, is a more appropriate measure of association for ordinal variables. Polychoric

correlations are based on the assumption that the observed variables are measuring continuous phenomena with only a few categories. Within this scenario, they are an unbiased estimator of the correlation between the underlying continuous variables (Babakus et al., 1987; Olsson, 1979a), and produce unbiased parameter estimates for both exploratory and confirmatory factor analysis (Babakus et al., 1987; Flora & Curran, 2004; Holgado-Tello, Chacón-Moscoso, Barbero-García, & Vila-Abad, 2010). However, polychoric correlations also have large sampling errors and frequently produce non-Gramian correlation matrices (indefinite matrices that have at least one negative eigenvalue), potentially compromising the effectiveness and applicability of the factor retention techniques (Tran & Formann, 2009; Weng & Cheng, 2005). Also, it should be noted that even though polychoric correlations are currently the standard measure of association for factor analysis with ordinal variables, some authors have questioned the appropriateness and desirability of the assumptions behind the polychorics, namely that the ordinal data comes from underlying continuous variables with bivariate normal distributions in the population (see Goodman & Kruskal, 1954).

Given the current state of the factor retention literature with ordinal variables, researchers and practitioners have no clear guidelines on how to proceed in determining the dimensionality of their ordinal data. In addition, one of the most accurate and recommended techniques for continuous variables (Henson & Roberts, 2006; Patil et al., 2008), the MAP method, has never been evaluated in the context of ordinal-level measurement. This study seeks to fill this void by taking an in-depth look at the performance of the MAP method with variables measured on only a few scale points.

3.1. Minimum Average Partial Method

The MAP method (Velicer, 1976) was developed in the context of principal component analysis, and is based on the matrix of partial correlations. Each component is partialled out

of the correlation matrix and the average of the squared partial correlations is computed. The number of factors to retain is determined by the point where the minimum average of the squared partial correlations is obtained. The rationale of this procedure can be described as follows: as common variance is partialled out of the correlation matrix for each successive component, the MAP criterion will keep on decreasing. At the point where the common variance has been removed, extracting additional components will result in unique variance being partialled out, and the MAP criterion will begin to rise. The MAP procedure, therefore, provides an unequivocal stopping point for the number of factors by separating the common and unique variance and retaining only those factors that consist primarily of common variance[1]. An illustration of the MAP method is presented in Figure 3.1.

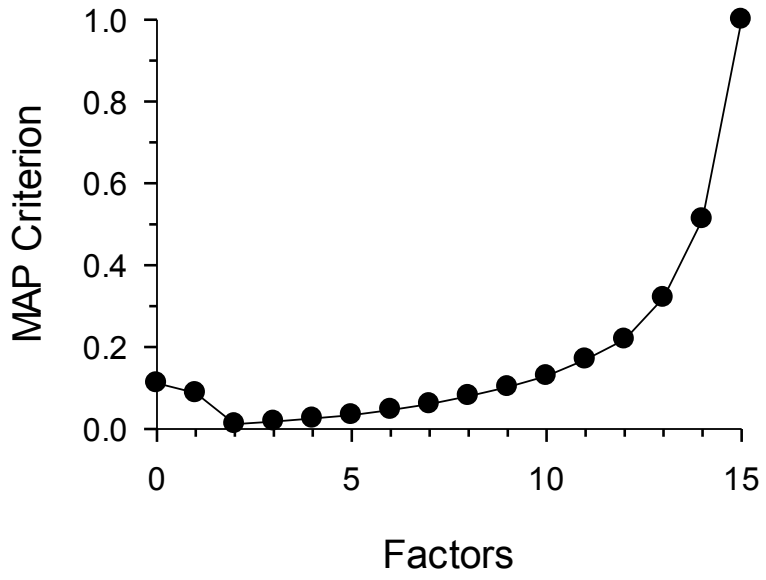


Figure 3.1: *An illustration of the MAP method*

The MAP procedure begins with the computation of the partial covariance matrix,

[1] Velicer et al. (2000) recommend the use of a component retention procedure (such as MAP or parallel analysis) even if the ultimate goal is a factor analysis. According to these authors, no satisfactory alternatives exist within factor analysis. The reader is directed to the works of Velicer and Jackson (1990) and Widaman (1993) for a review on the similarities and differences between principal component and factor analysis.

$$\mathrm{C}_m = \mathrm{R} - \mathrm{A}_m \mathrm{A}_m^{\mathrm{T}} \tag{3.1}$$

where C_m is the partial covariance matrix that results from partialling out the first m components from R, R is the correlation matrix, and A_m is the component loading matrix for components 1 to m.

Next, the partial correlation matrix is obtained

$$\mathrm{R}_m^* = \mathrm{D}^{-\frac{1}{2}} \mathrm{C}_m \mathrm{D}^{-\frac{1}{2}} \tag{3.2}$$

where R_m^* is the partial correlation matrix, and

$$\mathrm{D} = \mathrm{diag}(\mathrm{C}_m) \tag{3.3}$$

The MAP criterion is then obtained by averaging the squares of the partial correlations contained in R_m^*

$$\mathrm{MAP}_m = \sum_{\mathrm{i}=1}^{p} \sum_{\substack{\mathrm{j}=1 \\ \mathrm{i}\neq\mathrm{j}}}^{p} \frac{{\mathrm{r}_{\mathrm{ij}}^*}_m{}^2}{p(p-1)} \tag{3.4}$$

where p is the number of variables.

This procedure is repeated until $p-1$ components have been partialled out of R (partialling out p components would result in a null partial covariance matrix). Finally, Velicer proposes a test for the first factor by averaging the squares of the correlations contained in R

$$\mathrm{MAP}_0 = \sum_{\mathrm{i}=1}^{p} \sum_{\substack{\mathrm{j}=1 \\ \mathrm{i}\neq\mathrm{j}}}^{p} \frac{\mathrm{r}_{\mathrm{ij}}{}^2}{p(p-1)} \tag{3.5}$$

if $\mathrm{MAP}_0 < \mathrm{MAP}_1$, no factors should be extracted.

The efficacy of the MAP method with continuous variables has been evaluated in some of the most relevant factor retention studies to date. Zwick and Velicer (1982) carried out the first systematic examination of this procedure and found it to be more accurate than other stopping rules such as the eigenvalue-greater-than-1 rule and Bartlett's significance test (Bartlett, 1950, 1951). Their results also showed that it was affected primarily by the size of the factor loadings and had a tendency to underfactor with a small number of variables per factor. Using more complex data, Zwick and Velicer (1986) found the MAP procedure to be more accurate than the eigenvalue-greater-than-1 rule, Bartlett's test and the scree test, but inferior to parallel analysis. The same tendency to underfactor with low factor loadings and a small number of variables per factor was found in this study, replicating previous results. Sample size, on the other hand, had little effect on the accuracy of this method. Velicer et al. (2000) introduced a variant to the MAP procedure by raising the partial correlations to the 4^{th} power instead of squaring them. The results from their simulation study showed that this new variant produced slightly more accurate estimations than the original procedure. Finally, Peres-Neto et al. (2005) recommended the MAP method in their comparative study of 20 stopping rules, although the authors noted the tendency of the procedure to underfactor when it was in error.

3.2. Goals of the Current Study

The main goal of this study is to assess the performance of Velicer's MAP factor retention method with ordinal variables. We will focus on two areas of special interest: (1) determining the MAP variant that is the most effective; and (2) determining the factors that have the highest impact on the accuracy of the MAP estimations. We anticipate that some of the results with ordinal variables will be similar to those obtained with continuous data. Specifically, we expect that the factor loadings and the number of variables per factor will be the most salient independent variables. The factor correlation, on the other hand, has not been

manipulated in previous MAP studies (e.g., Zwick & Velicer, 1982, 1986; Velicer at al., 2000). In spite of this, we predict that it will be an important factor due to its impact on other factor retention methods (Crawford et al., 2010). Regarding the effectiveness of the different MAP variants, our theoretical expectation is that using polychoric correlations will produce more accurate estimations than using Pearson correlations, especially with skewed data. This is supported by the work on factor analysis with ordinal variables (Babakus et al., 1987; Olsson, 1979a). Nonetheless, some problems associated with the estimation of polychoric correlations may lead to unexpected results. In particular, MAP with polychorics may not perform well with smaller sample sizes or when the correlation matrices are non-Gramian. There is some evidence from studies with parallel analysis that suggest this possibility (Tran & Formann, 2009; Weng & Cheng, 2005). Finally, we have no *a priori* expectation regarding the effectiveness of raising the partial correlations to the 4^{th} power. Velicer et al. (2000) have been the only ones to study this variation, and they found it to be modestly superior to the original version of the method.

3.3. Method

3.3.1. Design

We employed a mixed factorial design to assess how effectively the MAP procedure estimates the correct number of factors. Two factors associated with the MAP method were manipulated: the type of correlation matrix to be analyzed, and the power that the partial correlations were raised to. Previous research suggests that the MAP method may be affected by these two factors (Babakus et al., 1987; Olsson, 1979b; Velicer et al., 2000). In addition, seven factors related to the data were systematically manipulated using Monte Carlo methods: the factor loading, number of variables per factor, number of factors, factor correlation, sample size, number of response categories, and skewness. These seven factors

have been shown to affect the performance of factor retention methods with continuous or ordinal variables (Cho et al., 2009; Weng & Cheng, 2005; Zwick & Velicer, 1982, 1986). A summary of the research design is presented in Table 3.1.

Table 3.1

Independent variables according to the research design

	Level					
Independent Variables	L1	L2	L3	L4	L5	L6
Method Factors						
Correlation Type	r	ρ				
MAP Power	2	4				
Data Factors						
Factor Loading	0.40	0.55	0.70			
Variables per Factor	4	8	12			
Number of Factors	2	4	6			
Factor Correlation	0.00	0.30	0.50			
Sample Size	200	500	1000			
Response Categories	2	3	4	5	6	7
Skewness	0	±1	±2			

Note. L = Level; r = Pearson; ρ = Polychoric.

Table 3.1 shows that the method factors make up a 2 x 2 design, producing 4 variants of the MAP method: $\mathbf{MAPr_{ij}^{*\,2}}$ (Pearson partial correlations squared), $\mathbf{MAPr_{ij}^{*\,4}}$ (Pearson partial correlations to the 4th power), $\mathbf{MAP\rho_{ij}^{*\,2}}$ (polychoric partial correlations squared), $\mathbf{MAP\rho_{ij}^{*\,4}}$ (polychoric partial correlations to the 4th power). On the other hand, the data factors make up a 3 x 3 x 3 x 3 x 3 x 6 x 3 design, resulting in a total of 4,374 factor combinations. The levels for the data factors were chosen so that they were representative of the range of values that are encountered in applied settings. In each case, an attempt was made to include a small/weak, medium/moderate, and large/strong level. For instance, according to Comrey & Lee (1992) factor loadings of 0.40, 0.55 and 0.70 can be considered as poor, good and excellent, respectively. Similarly, these authors consider sample sizes of 200, 500 and 1,000

to be fair, very good, and excellent. For the factor correlations, we included the orthogonal condition (r = 0.00), plus moderate (r = 0.30), and strong (r = 0.50) correlation levels, according to Cohen's (1988) criterion. Additionally, 4 variables per factor is just over the minimum of 3 that is required for factor identification (Widaman, 1993), 8 can be considered as a moderately strong factor (Velicer et al., 2000), and 12 as a highly overidentified factor (Widaman, 1993). Furthermore, we opted for a maximum of 7 response categories because the reliability and validity of ordinal measures appears to level off after reaching this number of scale points (Preston & Colman, 2000; Weng, 2004). The skewness of the ordinal variables was similarly varied from 0 (symmetric case) to ± 2, which according to Muthén and Kaplan (1985), borders on the limit of what is encountered with real data sets. Finally, the number of factors was varied from 2 to 6, representing relatively low and high values for modern inventories.

3.3.2. Data Generation

For each of the 4,374 factor combinations, 100 sample data matrices of ordinal variables were generated according to the following procedure: first, the reproduced population correlation matrix (with communalities in the diagonal) is computed

$$R_R = \boldsymbol{\Lambda\phi\Lambda}^T \tag{3.6}$$

where R_R is the reproduced population correlation matrix, $\boldsymbol{\Lambda}$ is the population factor loading matrix, and $\boldsymbol{\phi}$ is the population factor correlation matrix.

The population correlation matrix R_P is then obtained by inserting unities in the diagonal of R_R, thereby raising the matrix to full rank. The next step is performing a Cholesky decomposition of R_P, such that

$$R_P = U^T U \tag{3.7}$$

where U is an upper triangular matrix.

The sample matrix of continuous variables X is subsequently computed

$$X = ZU \tag{3.8}$$

where Z is a matrix of pseudo-random standard normal deviates with rows equal to the sample size and columns equal to the number of variables.

The sample matrix of ordinal variables is then obtained by applying a set of thresholds to X according to the specified levels of skewness and number of response categories (see Table 3.5 in the Appendix). The thresholds for the symmetric condition (skewness = 0) were computed by partitioning the continuum from z = -3 to z = 3 at equal intervals (see Bollen & Barb, 1981). Thresholds for the skewed conditions were created so that the observations were "piled up" in one of the extreme categories (see Muthén & Kaplan, 1985). In order to simulate "difficulty factors", half of the variables for each factor were categorized with positive skewness and the other half with negative skewness. An example of the frequency distribution of the ordinal variables used in this study is presented in Figure 3.2.

3.3.3. Assessment Criteria

The accuracy of the MAP method was evaluated according to three complementary criteria: the proportion of correct estimates (PC), the mean difference between the estimated and the correct number of factors (MD), and the root mean square error (RMSE). The corresponding formula for each criterion is presented in Equations 3.9 to 3.11:

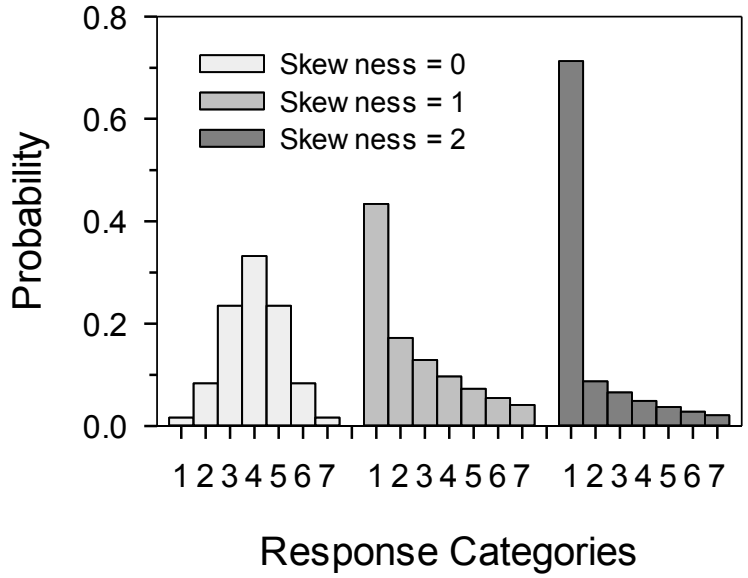


Figure 3.2: *Frequency distribution for the 7-point ordinal variables*

$$\mathrm{PC} = \frac{\sum \mathrm{C}}{\mathrm{N_S}}, \quad \text{for} \quad \mathrm{C} = \begin{cases} 1 & \text{if } \hat{m} = m \\ 0 & \text{if } \hat{m} \neq m \end{cases} \tag{3.9}$$

$$\mathrm{MD} = \frac{\sum(\hat{m} - m)}{\mathrm{N_S}} \tag{3.10}$$

$$\mathrm{RMSE} = \sqrt{\frac{\sum(\hat{m} - m)^2}{\mathrm{N_S}}} \tag{3.11}$$

where $\mathrm{N_S}$ is number of sample data matrices generated for each factor combination (100), $\hat{m}$ is the estimated number of factors, and m is the population number of factors.

The proportion of correct estimates criterion has boundaries of 0 and 1, with 0 indicating a total lack of accuracy and 1 reflecting perfect accuracy. In contrast, a 0 on the mean difference criterion shows a complete lack of bias, with negative and positive values indicating underfactoring and overfactoring, respectively. Similarly, higher RMSE values signal larger deviations from the population number of factors, while a value of 0 indicates

perfect accuracy. These three statistics were computed for each of the 4,374 factor combinations and were later averaged to obtain the values corresponding to each factor level.

3.3.4. Smoothing Procedure

Because polychoric correlations are usually estimated on a pairwise basis, that is, one correlation at a time, they often yield non-Gramian correlation matrices. Non-Gramian matrices have at least one eigenvalue that is negative and can't be inverted, something that is necessary in order to obtain the component-loading matrix needed for the computation of the MAP procedure. Given this situation, we decided to use a smoothing procedure to convert the non-Gramian matrices into Gramian matrices. The method we employed smoothens the non-Gramian matrix by adding a *ridge* (usually a small constant) to its diagonal until all the eigenvalues are positive (Wothke, 1993). At this point, the matrix is rescaled so that it becomes a correlation matrix once again:

$$R_{GR} = D^{-\frac{1}{2}} C_S D^{-\frac{1}{2}} \tag{3.12}$$

where R_{GR} is the smoothed correlation matrix, C_S is the covariance matrix that results from adding the ridge to the non-Gramian correlation matrix, and

$$D = \text{diag}(C_S) \tag{3.13}$$

After R_{GR} is obtained, the MAP procedure is computed exactly the same as was described previously. In order to analyze the possible impact of working with smoothed correlation matrices, these cases were flagged and studied separately.

3.4. Results

An overall assessment of the performance of the MAP method is presented in Table 3.2. Those cases with non-Gramian polychoric correlation matrices were first analyzed

separately, and then included in the global analysis. There were 85,130 non-Gramian polychoric matrices out of a total of 437,400 (19.46%). These cases occurred at a rate of 68.78% for N = 200, 24.16% for N = 500, and 7.06% for N = 1,000. Table 3.2 shows that the MAP methods with polychoric correlations were superior in every criterion to the MAP methods with Pearson correlations, even in those cases with non-Gramian polychoric matrices. Overall, the most effective method was $MAP\rho_{ij}^{*2}$, with more correct estimates (proportion of correct estimates = 0.59), less tendency to underfactor (mean difference = -1.30) and less error of estimation (RMSE = 1.37). Of special note is that the proportion of correct estimates for the polychoric methods, $MAP\rho_{ij}^{*2}$ and $MAP\rho_{ij}^{*4}$, is 0.20 and 0.25 higher than for its Pearson counterparts, $MAPr_{ij}^{*2}$ and $MAPr_{ij}^{*4}$.

Table 3.2

Overall MAP performance

	Non-Gramian PM			Total		
Method	PC	MD	RMSE	PC	MD	RMSE
$MAPr_{ij}^{*2}$	0.35	-1.46	2.23	0.39	-1.75	1.98
$MAPr_{ij}^{*4}$	0.22	-1.40	2.21	0.26	-1.76	2.07
$MAP\rho_{ij}^{*2}$	0.61	-1.35	1.45	0.59	-1.30	1.37
$MAP\rho_{ij}^{*4}$	0.55	-1.28	1.51	0.51	-1.34	1.46

Note. PM = Polychoric Matrix; PC = Proportion Correct; MD = Mean Difference; RMSE = Root Mean Square Error; r_{ij}^{*} = Pearson partial correlation; ρ_{ij}^{*} = Polychoric partial correlation.

Table 3.3 shows the performance for each number of response categories and for the original continuous variables (before they were categorized). The latter serve as a criterion for the maximum level of accuracy that can be expected with the ordinal variables. Several trends become apparent from Table 3.3: (1) the performance improves as the number of response categories increases, appearing to level-off at seven scale points; (2) polychoric

correlations produce more accurate estimations than Pearson correlations in all cases, with the differences becoming larger as the number of response categories decrease; (3) with many scale points, the performance of the polychoric based methods is close to the performance with the continuous variables, while the Pearson based methods appear to be biased (e.g., with 7 categories the RMSE of 1.36 for $MAP\rho_{ij}^{*4}$ is close to the 1.28 for the continuous variables, whereas the 1.89 for $MAPr_{ij}^{*4}$ is much larger); (4) $MAP\rho_{ij}^{*2}$ is the most accurate method for all the number of response categories that were studied (e.g., with 6 categories the mean difference of -1.20 for $MAP\rho_{ij}^{*2}$ is smaller than for any other method).

Table 3.3

MAP performance across the different number of response categories

	Number of Response Categories						
Method	2	3	4	5	6	7	Cont
				Proportion Correct			
$MAPr_{ij}^{*2}$	0.30	0.36	0.40	0.42	0.44	0.45	0.65
$MAPr_{ij}^{*4}$	0.20	0.24	0.27	0.28	0.30	0.31	0.55
$MAP\rho_{ij}^{*2}$	0.52	0.57	0.60	0.61	0.62	0.62	-
$MAP\rho_{ij}^{*4}$	0.47	0.50	0.52	0.53	0.53	0.53	-
				Mean Difference			
$MAPr_{ij}^{*2}$	-2.24	-1.92	-1.71	-1.61	-1.55	-1.49	-1.08
$MAPr_{ij}^{*4}$	-2.21	-1.89	-1.72	-1.63	-1.57	-1.53	-1.21
$MAP\rho_{ij}^{*2}$	-1.55	-1.37	-1.27	-1.23	-1.20	-1.19	-
$MAP\rho_{ij}^{*4}$	-1.56	-1.36	-1.31	-1.29	-1.27	-1.27	-
				Root Mean Square Error			
$MAPr_{ij}^{*2}$	2.35	2.12	1.95	1.88	1.83	1.78	1.13
$MAPr_{ij}^{*4}$	2.39	2.17	2.04	1.98	1.93	1.89	1.28
$MAP\rho_{ij}^{*2}$	1.63	1.45	1.34	1.30	1.27	1.25	-
$MAP\rho_{ij}^{*4}$	1.69	1.50	1.43	1.39	1.37	1.36	-

Note. Cont = Continuous; r_{ij}^{*} = Pearson partial correlation; ρ_{ij}^{*} = Polychoric partial correlation.

A detailed look at the performance of the MAP procedure for the rest of the factors is presented in Table 3.4. Some of the most notable findings are: (1) in accordance to the

theoretical expectations, all four MAP variants performed better with larger sample sizes, higher factor loadings, more variables per factor, less number of factors, lower factor correlation, and smaller skewness (a few exceptions occur for $MAPr_{ij}^{*4}$ and sample size); (2) on average, all variants of the MAP criterion underestimate the number of factors (all mean differences are negative); (3) the factor loadings and the number of variables per factor show the largest variability in accuracy for all criteria, therefore appearing to be the most relevant factors (e.g., with factor loadings of 0.40 the proportion of correct estimates for $MAP\rho_{ij}^{*2}$ is 0.23, much lower than with factor loadings of 0.70, where it is 0.93); (4) the superiority of the polychoric methods becomes more marked with larger levels of skewness; (5) $MAP\rho_{ij}^{*2}$ is, in general, the most accurate method according to every criterion.

Follow-up analyses were carried out to assess the performance of the Pearson methods in the conditions most likely to produce the difficulty factors. Specifically, the combination of highest factor loading (0.70), largest number of variables per factor (12), and greatest skewness (± 2) was studied. The results of these analyses revealed that the Pearson methods had a very low proportion of correct estimates (0.12 for $MAPr_{ij}^{*2}$ and 0.09 for $MAPr_{ij}^{*4}$) and overestimated the number of factors both on average (mean difference of 1.97 for $MAPr_{ij}^{*2}$ and 1.48 for $MAPr_{ij}^{*4}$) and in terms of total cases (89% of the cases had overestimations). Moreover, the rotated factor solutions showed that the additional factors suggested by the MAP procedure were made up of items with similar score distributions, a result that is consistent with the literature on difficulty factors. In contrast, the polychoric methods did not show the bias of the Pearson methods, as they did not overestimate the number of factors (mean difference of -0.03 for $MAP\rho_{ij}^{*2}$ and -0.01 for $MAP\rho_{ij}^{*4}$) and produced a very high proportion of correct estimates (0.98 for $MAP\rho_{ij}^{*2}$ and 0.93 for $MAP\rho_{ij}^{*4}$).

Table 3.4

MAP performance across the different levels of the independent variables

	Sample Size			Factor Loading			Variables per Factor			Number of Factors			Factor Correlation			Skewness		
Method	200	500	1000	0.40	0.55	0.70	4	8	12	2	4	6	0.00	0.30	0.50	0	±1	±2
									Proportion Correct									
$MAPr_{ij}^{*2}$	0.37	0.40	0.41	0.10	0.47	0.62	0.10	0.48	0.60	0.45	0.38	0.35	0.51	0.41	0.26	0.55	0.45	0.18
$MAPr_{ij}^{*4}$	0.30	0.26	0.24	0.06	0.35	0.38	0.04	0.29	0.46	0.38	0.23	0.18	0.35	0.26	0.18	0.45	0.27	0.08
$MAP\rho_{ij}^{*2}$	0.50	0.61	0.65	0.23	0.61	0.93	0.28	0.66	0.83	0.67	0.58	0.52	0.68	0.61	0.48	0.63	0.61	0.53
$MAP\rho_{ij}^{*4}$	0.45	0.54	0.55	0.17	0.55	0.82	0.18	0.58	0.78	0.64	0.48	0.42	0.63	0.52	0.39	0.54	0.53	0.47
									Mean Difference									
$MAPr_{ij}^{*2}$	-1.77	-1.75	-1.75	-3.01	-1.67	-0.59	-3.22	-1.53	-0.52	-0.79	-1.73	-2.75	-1.57	-1.68	-2.02	-1.45	-1.72	-2.10
$MAPr_{ij}^{*4}$	-1.73	-1.76	-1.79	-2.98	-1.74	-0.56	-3.15	-1.54	-0.59	-0.80	-1.77	-2.71	-1.43	-1.77	-2.07	-1.54	-1.69	-2.05
$MAP\rho_{ij}^{*2}$	-1.53	-1.24	-1.13	-2.43	-1.25	-0.23	-2.53	-0.95	-0.43	-0.50	-1.25	-2.16	-1.13	-1.22	-1.55	-1.15	-1.22	-1.54
$MAP\rho_{ij}^{*4}$	-1.49	-1.29	-1.25	-2.47	-1.16	-0.39	-2.52	-1.05	-0.47	-0.50	-1.35	-2.18	-1.00	-1.31	-1.72	-1.24	-1.28	-1.51
									Root Mean Square Error									
$MAPr_{ij}^{*2}$	2.02	1.97	1.96	3.05	1.72	1.17	3.25	1.63	1.07	0.93	1.96	3.06	1.82	1.92	2.21	1.50	1.81	2.64
$MAPr_{ij}^{*4}$	2.02	2.07	2.11	3.03	1.83	1.34	3.18	1.82	1.19	0.96	2.08	3.15	1.81	2.08	2.30	1.62	2.01	2.57
$MAP\rho_{ij}^{*2}$	1.64	1.31	1.18	2.51	1.30	0.31	2.61	1.03	0.49	0.55	1.32	2.25	1.21	1.28	1.64	1.21	1.29	1.62
$MAP\rho_{ij}^{*4}$	1.68	1.39	1.31	2.56	1.28	0.53	2.60	1.16	0.61	0.58	1.46	2.32	1.12	1.43	1.82	1.33	1.38	1.65

Note. r_{ij}^{*} = Pearson partial correlation; ρ_{ij}^{*} = Polychoric partial correlation.

The results shown in Tables 2, 3 and 4 revealed that the most accurate variant of the MAP method was the one with polychoric correlations and the square of the partial correlations (MAPρ_{ij}^{*2}). Consequently, we decided to conduct a more thorough analysis of the functioning of this method. Specifically, an analysis of variance (ANOVA) was performed with the MAPρ_{ij}^{*2} proportion of correct estimates as the dependent variable and the 7 data factors as the independent variables (using the mean difference and the RMSE as the dependent variables yielded very similar results). Due to the large sample size, most of the effects were significant. For this reason, we chose the partial eta squared (η_P^2) measure of effect size to establish the impact of the independent variables. According to Cohen (1988), η_P^2 values of 0.01 represent small effects, 0.06 medium effects, and 0.14 or more, large effects. Following this guide, the factor loading ($\eta_P^2 = 0.67$) and the variables per factor ($\eta_P^2 = 0.57$) had large effect sizes, and were by far the most salient variables. The factor correlation also had a large effect size ($\eta_P^2 = 0.15$), although it was considerably smaller than the ones for the two previous variables. In addition, medium effect sizes were found for the sample size ($\eta_P^2 = 0.09$) and the number of factors ($\eta_P^2 = 0.08$), while the skewness ($\eta_P^2 = 0.05$) and the number of response categories ($\eta_P^2 = 0.03$) produced small effects. Regarding the factor interactions, the factor loading x variables per factor interaction (see Figure 3.3) was the only one that reached a large effect size ($\eta_P^2 = 0.38$); the other less salient interactions did not appear to have theoretical or practical implications and therefore will not be discussed here.

The most notable feature of the factor loading x variables per factor interaction is that with 4 variables per factor there is a negligible improvement in accuracy when the factor loadings increase from 0.40 to 0.55; the improvement only comes when the loadings increase from 0.55 to 0.70. In contrast, there are substantial increases in the proportion of correct

estimates for 8 (from 0.15 to 0.85) and 12 (from 0.54 to 0.95) variables per factor. Also of note is that the difference in accuracy between 8 and 12 variables per factor is greatly reduced as the loadings increase to 0.55 (the difference diminishes from 0.39 to 0.10).

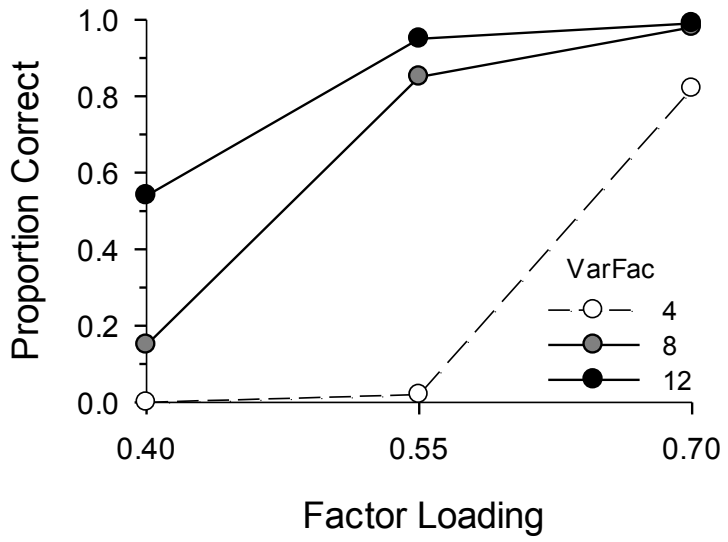


Figure 3.3: *Factor Loading x Variables per Factor (VarFac) interaction*

In order to better explain the factor loading x variables per factor interaction, we conducted an additional simulation at the population level. Each of the 4 relevant population factors was manipulated: factor loading, number of variables per factor, number of factors, and factor correlation. In this case, we chose 3 levels of factor loading (0.40, 0.55 and 0.70), 13 levels of variables per factor (from 3 to 15), 10 levels of number of factors (from 1 to 10), and 6 levels of factor correlations (from 0.00 to 0.50 in increments of 0.10). A summary of the results from this simulation study across the levels of factor loading and variables per factor is presented in Figure 3.4.

The results in Figure 3.4 show that the MAP method has a clear bias in two conditions: for low factor loadings (0.40) and for medium factor loadings (0.55) and a small number of variables per factor (≤ 6). In these cases, the MAP method doesn't always return

the correct number of factors, something that would be expected since the data is free from sampling error. Of special note is that the erroneous MAP estimates occur in exactly the same cases where the component loadings are known to have a large upward bias, that is, with low communalities and/or a small number of variables per factor (Widaman, 1993). The problem can be explained with the help of Equation 3.1: an upward bias in the component loadings results in an upward bias in the reproduced correlations, and consequently, in large partial covariances/correlations (the partial covariances are obtained by subtracting the reproduced correlations from the original correlations). In turn, these large partial correlations produce an increase in the MAP criterion, and the factor is discarded. This is why the MAP method has always shown a tendency to underfactor, because it will discard the factors with low communalities and/or a small number of variables per factor, even at the population level.

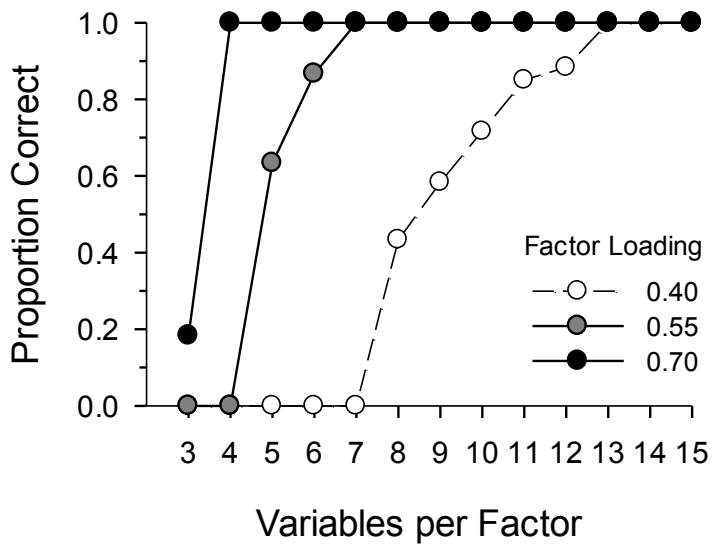


Figure 3.4: *MAP performance at the population level*

3.5. Discussion

The current study used Monte Carlo methods to assess the performance of Velicer's MAP factor retention method with ordinal variables, typically encountered in the educational

and psychological fields. The MAP procedure has been identified through numerous studies as one of the most accurate techniques to assess data dimensionality (Peres-Neto et al., 2005; Velicer et al., 2000; Zwick & Velicer, 1986), a critical phase of an exploratory factor analysis (EFA). In addition, this procedure is currently recommended in "best practice" EFA guidelines (e.g., Hayton et al., 2004; Henson & Roberts, 2006; Patil et al., 2008), and its use has been increasing in recent years due to the availability of easy-to-use computer programs (e.g., O'Connor, 2000). However, despite this widespread support, until now no study had assessed its efficacy with ordinal-level measurement, even though previous research has shown that methods developed for continuous variables may produce biased results if used with ordinal data (Babakus et al., 1987; Olsson, 1979b).

In order to perform a comprehensive evaluation of the MAP procedure, we employed a mixed factorial design that included the systematic manipulation of a set of factors related to the MAP method and to the data. In terms of the *method factors*, the results indicated that the polychoric correlations lead to considerably more accurate dimensionality estimations than the Pearson correlations, especially in the presence of skewed data. These findings are consistent with the factor analysis literature, which shows that that the polychoric correlation coefficient is a more appropriate measure of association for ordinal variables (Olsson, 1979a), and that Pearson correlations are especially biased for skewed ordinal data (Babakus et al., 1987; Olsson, 1979b). In addition, the estimates derived from polychoric correlations were more accurate for both Gramian and smoothed non-Gramian polychoric matrices, alleviating concerns that have been raised in previous studies with polychoric correlations (e.g., Tran & Formann, 2009). As far as the power that the partial correlations are raised to, the results showed superior performance for the squared partial correlations in nearly all the conditions studied. Therefore, raising the partial correlations to the 4^{th} power is not recommended.

Moving on to the *data factors*, an in-depth analysis was carried out for the most accurate version of the MAP procedure, the one with polychoric correlations and the square of the partial correlations. The results from the corresponding analysis of variance (ANOVA) indicated that the factor loadings and the number of variables per factor have by far the highest impact on the accuracy of the MAP method; they are followed by the factor correlation (large effect size), the sample size and the number of factors (moderate effect sizes), and the skewness and the number of response categories (small effect sizes). In addition, the MAP method showed a strong tendency to underestimate the number of factors when it was in error. These results were expected and fully replicate the behavior of the MAP method with continuous variables, where the factor loadings and the number of variables per factor have been the most salient factors (Zwick & Velicer, 1982, 1986), the sample size has exhibited minimum impact (Zwick & Velicer, 1986), and the tendency has been to underestimate the number of factors (Peres-Neto et al., 2005). The factor correlation, which had not been included in previous MAP research, had an important effect on the accuracy of the estimations, a result that is consistent with the findings pertaining to other factor retention rules (Crawford et al., 2010).

A novel finding of this study is the identification of the factor loading x number of variables per factor interaction as having a critical role in the effectiveness of the MAP procedure. The main feature of this interaction is that with 4 variables per factor there is negligible improvement in accuracy when the factor loadings are raised from 0.40 to 0.55, while with 8 or 12 variables per factor the increase is dramatic. A secondary feature of the interaction is that the benefit of having 12 variables per factor over 8 is reduced substantially with medium (0.55) or higher loadings. This latter result may guide researchers when weighing the cost/benefits of having more than 8 variables per factor as it pertains to recovering the true dimensionality of the data. Regarding the main feature of the interaction,

the finding was surprising and as a result we decided to conduct a follow-up simulation study at the population level. Initially, it had been assumed that the MAP procedure would yield perfect estimates at the population level, where the correlation matrices are free from sampling error. However, this new simulation revealed that the MAP method produces incorrect estimates for two conditions: (1) for low factor loadings (0.40); and (2) for medium factor loadings (0.55) and a small number of variables per factor (≤ 6). The reason for this bias (and for the main feature of this interaction) is that with low communalities and/or a small number of variables per factor the component loadings have a substantial upward bias, which in turn results in upwardly biased reproduced correlations and in large partial correlations. These large partial correlations are what cause the MAP criterion to increase and these factors to be discarded. This also explains why the MAP method has always shown a tendency to underestimate the number of factors, because even at the population level, it will tend to discard factors with the abovementioned characteristics.

Taking into account the combined results for the method and data factors, as well as the simulation at the population level, we propose the following guidelines to researchers who wish to use this factor retention method with ordinal variables:

1. Use polychoric correlations and the square of the partial correlations.
2. Smooth the non-Gramian polychoric matrices with the ridge procedure.
3. Design scales to be ultimately defined by at least 6 variables.

There are some limitations in this study that should be noted. First, all the models had perfect simple structure, and the factor loadings, number of variables per factor, and factor correlations were all homogeneous within cases. Since these idealized models are unlikely to be encountered in practice, the results of this study should be viewed as a best-case scenario. Second, there was only one factor retention method studied. The reason for this focused

approach was the lack of previous research from which to build upon. However, future studies should test this procedure alongside other methods to determine its relative level of effectiveness.

This study has provided a detailed look at the performance of the MAP method with ordinal variables. Until now, researchers who wished to use this procedure to determine the dimensionality of their ordinal data have had to rely on the results and recommendations from studies with continuous variables, and in many cases, have had to employ programs that were developed exclusively for this level measurement (e.g., Eklöf, 2006; Lai, Crane, & Cella, 2006; Wood, Maltby, & Joseph, 2008). We hope this study can serve as a first step in changing this practice, by helping researchers to make more informed decisions and encouraging the creation of new MAP computer programs adapted for ordinal variables.

3.6. Appendix

Table 3.5

Thresholds used to obtain the ordinal variables

	Thresholds					
# RC	1	2	3	4	5	6
	Skewness = 0					
2	0.00					
3	-1.00	1.00				
4	-1.50	0.00	1.50			
5	-1.80	-0.60	0.60	1.80		
6	-2.00	-1.00	0.00	1.00	2.00	
7	-2.14	-1.29	-0.43	0.43	1.29	2.14
	Skewness = 1					
2	0.59					
3	0.32	0.99				
4	0.17	0.69	1.25			
5	0.05	0.51	0.94	1.45		
6	-0.06	0.38	0.76	1.14	1.61	
7	-0.17	0.27	0.63	0.96	1.31	1.74
	Skewness = 2					
2	1.05					
3	0.85	1.38				
4	0.75	1.13	1.60			
5	0.68	1.00	1.34	1.77		
6	0.62	0.91	1.20	1.51	1.91	
7	0.56	0.84	1.11	1.37	1.66	2.04

Note. # RC = Number of Response Categories; For negative skewness change the sign of the thresholds.

4. STUDY 2: A NEW LOOK AT HORN'S PARALLEL ANALYSIS WITH ORDINAL VARIABLES

Previous research evaluating the performance of Horn's parallel analysis (PA) factor retention method with ordinal variables has produced unexpected findings. Specifically, PA with Pearson correlations has performed as well or better than PA with the more theoretically appropriate polychoric correlations. Seeking to clarify these findings, the current study employed a more comprehensive simulation study that included the systematic manipulation of seven factors related to the data (sample size, factor loading, number of variables per factor, number of factors, factor correlation, number of response categories and skewness) as well as three factors related to the PA method (type of correlation matrix, extraction method, and eigenvalue percentile). The results from the simulation study show that PA with either Pearson or polychoric correlations is particularly sensitive to the sample size, factor loadings, number of variables per factor and factor correlations. However, whereas PA with polychorics is relatively robust to the skewness of the variables, PA with Pearson correlations frequently retains difficulty factors and is generally inaccurate with large levels of skewness. In light of these findings, we recommend the use of PA with polychoric correlations for the dimensionality assessment of ordinal-level data.

One of the primary uses of exploratory factor analysis (EFA) in the educational and psychological fields is to identify the underlying dimensions of a domain of functioning, as assessed by a particular measuring instrument (Floyd & Widaman, 1995). A key decision in this process is determining the number of factors to retain for a group of variables of interest (Fabrigar, Wegener, MacCallum, & Strahan, 1999; Hayton, Allen, & Scarpello, 2004; Henson & Roberts, 2006; Velicer, Eaton, & Fava, 2000). This decision is especially

important because errors of underfactoring (extracting too few factors) or overfactoring (extracting too many factors) are likely to result in non-interpretable or unreliable factors (Fava & Velicer, 1992, 1996; Lee & Comrey, 1979; Wood, Tataryn, & Gorsuch, 1996), and can potentially mislead theory development efforts (Fabrigar et al., 1999; Patil, Singh, Mishra, & Donovan, 2008).

There are many rules for deciding the number of factors to retain, and they can be divided into three categories (Floyd & Widaman, 1995): statistical tests, mathematical and psychometric criteria and rules of thumb. *Statistical tests* for EFA are available for some estimation methods such as maximum likelihood, generalized least squares or asymptotically distribution free methods, and are computed as chi-square significance tests of the residual covariation among observed variables after extracting a certain number of factors. The *mathematical and psychometric* criteria includes some of the most widely used and/or recommended methods such as the eigenvalue-greater-than-1 rule or Kaiser-Guttman criterion (Kaiser, 1960), parallel analysis (Horn, 1965), and the minimum average partial method (Velicer, 1976). Of these, the eigenvalue-greater-than-1 rule constitutes the default in most statistical packages and is based on population proofs regarding the size of the eigenvalues for uncorrelated variables. In addition, many practical criteria falls under the rubric of *rules of thumb*, like the scree test (Cattell, 1966), the percentage of variance accounted for, and the ratio of first-to-second eigenvalues.

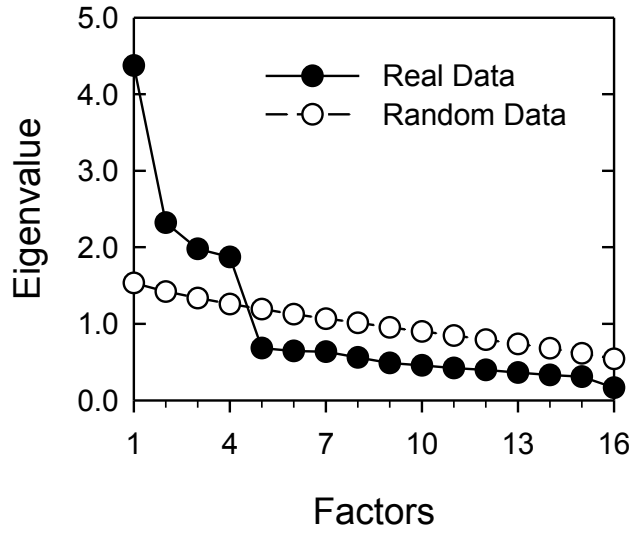
Among the many factor retention methods that have been proposed, Horn's parallel analysis (**PA**) has emerged as one of the most accurate and recommended dimensionality assessment techniques for continuous data (Fabrigar et al., 1999; Hayton et al., 2004; Henson & Roberts, 2006; Peres-Neto, Jackson, & Somers, 2005; Velicer et al., 2000; Zwick & Velicer, 1986). In addition, the use of PA has been increasing in the past decade due to the development of syntax code for some of the most popular statistical packages like SPSS, SAS

and Stata (Dinno, 2009; O'Connor, 2000), and to the inclusion of the procedure in factor analysis software such as FACTOR (Lorenzo-Seva & Ferrando, 2006). Also, the interest in PA has extended in recent years to the dimensional assessment of ordinal variables, which are commonly found in datasets from self-report and achievement tests. In this area, PA has quickly become the most studied factor retention method (see Cho, Li & Bandalos, 2009; Timmerman & Lorenzo-Seva, 2011; Tran & Formann, 2009; Weng & Cheng, 2005).

4.1. Parallel Analysis Method

Horn (1965) proposed the PA method on the basis of Kaiser's (1960) and Dickman's (1960) proofs and arguments that Guttman's (1954) latent-root-one lower bound estimate for the minimum rank of a correlation matrix could be used as a psychometric upper bound for the number of factors problem. The eigenvalue-greater-than-1 criterion, also known as K1 or Kaiser's rule, posits that only factors with eigenvalues > 1 should be retained. Part of the rationale of this rule is that a factor should be able to explain at least as much variance as a variable is accorded in the standard score space (Dickman, 1960), and that a threshold of 1 ensures that the component will have a positive internal consistency (Kaiser, 1960). Because the proofs for the eigenvalue-greater-than-1 rule were performed on population statistics, Horn (1965) argued that due to sampling error, and least squares "capitalization" on this error in the computation of the latent roots, some components from uncorrelated variables in the population could have eigenvalues > 1 at the sample level. Therefore, he proposed PA as a means to estimate and take into account the proportion of variance that was due to sampling error and chance capitalization. In this sense, PA may be viewed as a sample alternative to the eigenvalue-greater-than-1 rule. Instead of retaining factors that have eigenvalues > 1, with PA only those factors that have eigenvalues greater than those generated from independent variates are retained. The goal is to account for chance capitalization in the sample eigenvalues under the null hypothesis of independent variables (Buja & Eyuboglu, 1992).

The implementation of the PA procedure involves the generation of a large number of matrices of random data. Each matrix is generated with the same number of subjects and variables as the real data matrix under assessment. Then, the number of factors is determined by comparing the eigenvalues from the real data matrix with the mean of the eigenvalues from the random data matrices (Horn, 1965). A factor is retained as long as its eigenvalue is greater than the mean eigenvalue from its random counterpart. An example of the PA method is presented in Figure 4.1. In this case, a 4-factor solution is suggested.



Note. The values for the random data are the mean eigenvalues across 100 samples

Figure 4.1: *An illustration of the Parallel Analysis Method*

A number of modifications of the PA procedure have been suggested over the years. One group of modifications is related to the extraction method used to compute the eigenvalues. Instead of using Principal Component Analysis (**PA-PCA**), as in Horn's (1965) original formulation, some authors have suggested the use of extraction methods that fit the common factor model. In this line, Humphreys and Ilgen (1969) proposed a PA variant with Principal Axes Factor Analysis (**PA-PAFA**), where the eigenvalues are computed from a reduced correlation matrix with squared multiple correlations in the diagonal. This approach,

however, appears to be theoretically inappropriate because the squared multiple correlation is a biased estimate of the true communality of a variable (Buja & Eyuboglu, 1992). In order to overcome the limitations of PA-PAFA, Timmerman and Lorenzo-Seva (2011) recently proposed a variant of PA with Minimum Rank Factor Analysis (**PA-MRFA**). In an initial study with ordinal variables, they found PA-MRFA to be marginally superior to PA-PCA, and substantially more accurate than PA-PAFA.

Another group of modifications deals with the aggregation rule for the random eigenvalues. Horn (1965) originally proposed the mean of the K sets of random eigenvalues as the criterion (**PA_m**). According to early PA research (e.g., Zwick & Velicer, 1986), the mean criteria tended to overextract the numbers of factors; as a result, some researchers suggested that more stringent criteria based on inferential theory should be used, like the 95th percentile (**PA_{95}**; Glorfeld, 1995; Weng & Cheng, 2005). In this sense, one might interpret PA as a method to assess the significance of each factor (Glorfeld, 1995), although this is only appropriate for the first eigenvalue due to the inherent dependencies between successive eigenvalues (Buja & Eyuboglu, 1992). Further research in this area with more complex designs have shown that the 95th percentile works better for a single factor or uncorrelated factors, while the mean rule is more accurate for correlated structures (Cho et al., 2009; Crawford et al., 2010).

4.2. Dimensionality Assessment of Ordinal Variables with Parallel Analysis

Several factors make the dimensionality assessment of ordinal data more difficult than for normally distributed continuous variables. As it is well known, Pearson's product-moment correlation underestimates the strength of the relationship between ordinal variables (Babakus, Ferguson, & Jöreskog, 1987; Bollen & Barb, 1981), and may produce spurious dimensions known as "difficulty factors" when the variables are skewed in opposite

directions (Gorsuch, 1983; Olsson, 1979b). Because of these biases, the polychoric correlation coefficient has been recommended as a measure of association for the factor analysis of ordinal variables (Flora & Curran, 2004; Jöreskog & Moustaki, 2001). Assuming that the ordinal variables are a crude measure of underlying bivariate normally distributed variables, the polychoric correlation is a maximum likelihood estimate of the Pearson correlation between the underlying variables (Olsson, 1979a). Polychoric correlations have been shown to produce unbiased parameter estimates for both exploratory and confirmatory factor analysis (Babakus et al., 1987; Flora & Curran, 2004). Despite these advantages, however, polychoric correlations have some problems of their own. In particular, they frequently produce non-Gramian correlation matrices (indefinite matrices that have at least one negative eigenvalue), have large sampling errors, and can take considerable time to estimate, properties that can potentially compromise the effectiveness and applicability of factor retention methods such as PA (Timmerman & Lorenzo-Seva, 2011; Tran & Formann, 2009; Weng & Cheng, 2005).

Support for the dimensionality assessment of ordinal data with PA and the more appropriate polychoric correlations has so far been limited. Weng and Cheng (2005) compared empirical eigenvalues from Pearson and tetrachoric correlations with random eigenvalues from a multivariate normal distribution in order to assess the effectiveness of PA with unidimensional binary data. The results from their simulation study with positively skewed variables showed that PA with Pearson correlations (**PAr**) was more accurate than PA with polychorics (**Paρ**), a finding they attributed to the large sampling errors and unstable behavior of the tetrachoric correlations. A subsequent study by Tran and Formann (2009) extended the evaluation of PA with unidimensional binary data by simulating factors with both positively and negatively skewed items, the scenario most likely to produce difficulty factors with Pearson correlations. The results from their study indicated that neither PAr nor

Paρ could be recommended; in the case of PAr, because of poor performance, and in the case of Paρ, due to applicability issues stemming from the large number of non-Gramian matrices.

Cho et al. (2009) further advanced the study of PA with ordinal variables by assessing its performance with polytomous items of 2 and 3 response options and multidimensional structures of symmetrically distributed variables. In this case, the authors matched the type of correlation matrix used to compute the real and random eigenvalues and found that PAr was at least as accurate as Paρ. More recently, Timmerman & Lorenzo-Seva (2011) studied the effectiveness of PA with multidimensional structures of polytomous items skewed in opposite directions. According to their results, Paρ could only be computed for 37% of the data matrices due to convergence problems of the algorithm used to compute the polychoric correlations or because the smoothing procedure did not produce a Gramian polychoric matrix. This situation prompted the authors to state that the "convergence problems of the polychoric approach prevent its general application to empirical data" and "may pose severe problems" in practice (p. 218).

As can be seen from the preceding commentary, the performance of Paρ has been extremely difficult to gauge due to the problems associated with non-Gramian polychoric matrices. This issue, coupled with the incomplete or non-manipulation of the key factor of skewness in some studies, is currently making the comparison of PA with Pearson and polychoric correlations particularly challenging. It appears, therefore, that a more accurate representation of the behavior of PA with ordinal variables may be obtained from a new study that addresses these issues simultaneously.

4.3. Proposal for the Current Study

The proposal of the current study is to assess the efficacy of PA with the inclusion of all polychoric matrices, Gramian and non-Gramian, in the context of a simulation study that systematically manipulates the skewness of the ordinal variables as well as the other relevant factors. The performance of PA with non-Gramian matrices may be tested using two approaches: (1) smoothing the matrices in order to eliminate the negative eigenvalues, and (2) using the eigenvalues as they are without any treatment. The theoretical and practical implications of both approaches are discussed below.

Although proper factor solutions can be obtained from non-Gramian polychoric matrices (Babakus et al., 1987; Flora & Curran, 2004), they make PA difficult to interpret because the eigenvalues are no longer related to the variance explained of a factor (due to some eigenvalues being negative). Smoothing the non-Gramian matrices with a procedure that always produces a Gramian matrix (e.g., the Eigenvalue method; Knol & ten Berge, 1991), however, may resolve this problem. This approach would maintain the precise rationale of the PA method as all eigenvalues would now be positive and its potential impact could be tested empirically. An alternative approach also worth considering would be to apply PA to the unsmoothed non-Gramian matrices, using all positive and negative eigenvalues as they are. Even though under this methodology the rationale of PA may not be theoretically straightforward for the reasons outlined above, the practical implications may be minimal. Horn (1965) proposed PA as a means of "subtracting out the component in the latent roots which can be attributed to sampling error, and least-squares "capitalization" on this error, in the calculation of the correlations and the roots" (p. 179). In this sense, if the random data is modeled closely to the real data (e.g., via random column permutation of the real data), the sampling error in the polychoric correlations that produces the non-Gramian matrices may work similarly for both types of data, making the comparison of their

eigenvalues more or less unbiased or unaffected. Under this scenario PA with non-Gramian matrices would still maintain the general rationale outlined by Horn: the procedure would be estimating and taking into account the amount in the eigenvalues that is due to sampling error and chance capitalization.

4.4. Goals of the Current Study

The main goal of the present study is to compare the effectiveness of PA with Pearson and polychoric correlations in determining the number of common factors present in datasets of ordinal variables. A secondary goal is to determine the impact of a comprehensive set of factors, and their interactions, in the accuracy of the different variations of the PA procedure.

Regarding the main goal of this study, the performance of PAr and Paρ is expected to be similar with symmetrical distributions, as in Cho et al. (2009). The reason that PAr should work well with unskewed data is that the Pearson correlations are likely to be underestimated similarly for the real and random data (assuming that the random data is modeled closely to the real data), therefore limiting the potential bias in the comparison of the eigenvalues. With skewed data, however, Paρ should outperform PAr due to the emergence of difficulty factors with the Pearson correlations (Gorsuch, 1983; Olsson, 1979b). Also, the superiority of Paρ over PAr should be more evident with skewed data that has high factor loadings, as the bias introduced by the difficulty factors becomes more salient in this condition (Olsson, 1979b). In terms of the secondary goal of this study, prior research with ordinal and continuous data suggests that the factor loadings, sample size, number of factors and factor correlations are all important variables that affect the accuracy of PA (Beauducel, 2001; Cho et al., 2009; Weng & Cheng, 2005; Zwick & Velicer, 1986). They are expected to be salient factors here as well. In addition, using the mean of the random eigenvalues is expected to produce more accurate estimations with correlated factors (Cho et al., 2009), whereas the 95th percentile criterion should work better for uncorrelated structures (Crawford et al., 2010).

Finally, early research of PA with MRFA extraction has shown it to be marginally superior to PA with PCA extraction. Because PA-MRFA is a common factor model variation of the PA method, it may offer a more accurate estimation of the number common factors present in the data. On the other hand, PA-PCA may be robust to the known biases of PCA, namely, the overestimation of the variable saturation when the population loadings are low and/or the number of variables per factor is small. The reason that PA may be robust to the biases of PCA extraction is, again, the fact that both the real and random data will be affected similarly, thus rendering their comparison relatively unbiased. One of the aims of this research is to be able to answer this question satisfactorily.

4.5. Method

4.5.1. Design

A mixed factorial design was employed to assess the effectiveness of the different PA methods. Three within-subjects "method" factors were manipulated: the type of correlation matrix used to compute the eigenvalues (Pearson or polychoric), the extraction method (PCA or MRFA), and the percentile of the random eigenvalues (the mean or the 95th percentile). In addition, seven between-subject "data" factors were systematically manipulated using Monte Carlo methods: the sample size, factor loading, number of variables per factor, number of factors, factor correlation, number of response categories, and skewness. Altogether, these 10 factors have been shown to affect the performance of factor retention methods with ordinal and/or continuous variables (Cho et al., 2009; Timmerman & Lorenzo-Seva, 2011; Tran & Formann, 2009; Velicer et al., 2000; Weng & Cheng, 2005; Zwick & Velicer, 1986). A summary of the research design is presented in Table 4.1.

Table 4.1 shows a 2 x 2 x 2 (correlation type x extraction method x eigenvalue percentile) within-subjects design that produces 8 variants of the PA method: (1) **PA-PCA$\mathbf{r_m}$**;

(2) **PA-PCAr$_{95}$**; (3) **PA-PCAρ$_{m}$**; (4) **PA-PCAρ$_{95}$**; (5) **PA-MRFAr$_{m}$**; (6) **PA-MRFAr$_{95}$**; (7) **PA-MRFAρ$_{m}$**; and (8) **PA-MRFAρ$_{95}$**; where PA = Parallel Analysis; PCA = Principal Component Analysis; MRFA = Minimum Rank Factor Analysis; r = Pearson correlations; ρ = polychoric correlations; m = mean eigenvalue criterion; and 95 = 95th percentile eigenvalue criterion. In terms of the between-subject factors, the design can be divided into two parts: (1) the unidimensional condition with a 3 x 3 x 3 x 4 x 5 (sample size x factor loading x variables per factor x response categories x skewness) design, for a total of 540 factor combinations; and (2) the multidimensional condition with a 3 x 3 x 3 x 3 x 3 x 4 x 5 (sample size x factor loading x variables per factor x number of factors x factor correlation x response categories x skewness) design, for a total of 4,860 factor combinations. In all, 5,400 between-subjects factor combinations were studied.

Table 4.1

Independent Variables According to the Research Design

	Level				
Independent Variables	L1	L2	L3	L4	L5
Method Factors					
Correlation Type	r	ρ			
Extraction method	PCA	MRFA			
Eigenvalue Percentile	m	95			
Data Factors					
Sample Size	100	300	1000		
Factor Loading	0.40	0.55	0.70		
Variables per Factor	4	8	12		
Number of Factors	1	2	4	6	
Factor Correlation	0.00	0.30	0.50		
Response Categories	2	3	4	5	
Skewness	0.00	±0.50	±1.00	±1.50	±2.00

Note. L = level; PCA = principal component analysis; MRFA = minimum rank factor analysis; r = Pearson; ρ = polychoric; m = mean eigenvalue; 95 = 95% eigenvalue.

The levels for the data factors were chosen so that they were representative of the range of values that are encountered in applied settings. In each case, an attempt was made to

include a small/weak, medium/moderate, and large/strong level. For instance, according to Comrey and Lee (1992) sample sizes of 100, 300 and 1,000 can be considered as poor, good, and excellent. Similarly, these authors consider factor loadings of 0.40, 0.55 and 0.70 to be poor, good and excellent as well. For the factor correlations, the orthogonal condition (r = 0.00) was included, plus moderate (r = 0.30) and strong (r = 0.50) correlation levels, according to Cohen's (1988) criterion. Additionally, 4 variables per factor is just over the minimum of 3 that is required for factor identification (Widaman, 1993), 8 can be considered as a moderately strong factor (Velicer et al., 2000), and 12 as a highly overidentified factor (Widaman, 1993). Furthermore, 5 was chosen as the maximum number of response categories to be simulated because gains in reliability and validity appear to be only marginal with more scale points (Preston & Colman, 2000). In terms of the skewness of the ordinal variables, they were varied from 0.00 to ±2.00 in increments of ±0.50. A skewness level of 0.00 indicates a symmetrical distribution, whereas ±1.00 may be considered as a meaningful departure from normality (Meyers, Gamst, & Guarino, 2006, p. 50), and ±2.00 as a high level of skewness that borders on the limit of what is typically encountered with real data sets (Muthén & Kaplan, 1985). The lower levels of skewness may be characteristic of attitude tests and personality inventories, while the larger levels of oppositely skewed variables may be found on aptitude tests (such as intelligence batteries) where the items are designed to have difficulty levels that range from very easy to very difficult. Finally, the number of factors was varied from 1 to 6, which includes the unidimensional condition, as well as relatively low to high values for modern multidimensional inventories.

4.5.2. Data Generation

For each of the 5,400 factor combinations, 100 sample data matrices of ordinal variables were generated according to the following common factor model procedure: first,

the reproduced population correlation matrix (with communalities in the diagonal) is computed

$$\mathbf{R}_{\mathrm{R}} = \boldsymbol{\Lambda}\boldsymbol{\phi}\boldsymbol{\Lambda}^{\mathrm{T}} \tag{4.1}$$

where $\mathbf{R}_{\mathrm{R}}$ is the reproduced population correlation matrix, $\boldsymbol{\Lambda}$ is the population factor loading matrix, and $\boldsymbol{\phi}$ is the population factor correlation matrix.

The population correlation matrix $\mathbf{R}_{\mathrm{P}}$ is then obtained by inserting unities in the diagonal of $\mathbf{R}_{\mathrm{R}}$, thereby raising the matrix to full rank. The next step is performing a Cholesky decomposition of $\mathbf{R}_{\mathrm{P}}$, such that

$$\mathbf{R}_{\mathrm{P}} = \mathbf{U}^{\mathrm{T}}\mathbf{U} \tag{4.2}$$

where $\mathbf{U}$ is an upper triangular matrix.

The sample matrix of continuous variables $\mathbf{X}$ is subsequently computed

$$\mathbf{X} = \mathbf{Z}\mathbf{U} \tag{4.3}$$

where $\mathbf{Z}$ is a matrix of random standard normal deviates with rows equal to the sample size and columns equal to the number of variables.

The sample matrix of ordinal variables is obtained by applying a set of thresholds to $\mathbf{X}$ according to the specified levels of skewness and number of response categories (see Table 4.7 in the Appendix). The thresholds for the symmetric condition (skewness = 0.00) were computed by partitioning the continuum from z = -3 to z = 3 at equal intervals (see Bollen & Barb, 1981). Thresholds for the asymmetric conditions were created so that as the skewness level increased, the observations were “piled up” in one of the extreme categories (see Muthén & Kaplan, 1985). In order to simulate “difficulty factors”, half of the variables for

each factor were categorized with the same positive skewness and the other half with the same negative skewness.

In order to obtain the criterion eigenvalues for the PA methods, 100 random data matrices of standard normal deviates were generated for each combination of sample size and number of variables. This number of replicates is considered to be sufficient to yield stable criterion eigenvalues (Buja & Eyuboglu, 1992). Next, the same set of thresholds used to obtain the "real" ordinal data was applied to the random data generated in the previous step. After computing the eigenvalues for each of the 100 correlation matrices, they were combined according to the mean and the 95th percentile criteria.

4.5.3. Smoothing Procedure

The non-Gramian polychoric correlation matrices were smoothed using the Eigenvalue method described in Knol and ten Berge (1991). This method is based on an eigendecomposition of the improper correlation matrix, followed by the replacement of the negative eigenvalues with a small positive constant and the computation and rescaling of the covariance matrix with the new eigenvalues. A constant of +0.01 was used in this study in order to ensure that the matrix to be analyzed was sufficiently well conditioned.

4.5.4. Assessment Criteria

The accuracy of the PA methods was evaluated according to three complementary criteria: the proportion of correct estimates (PC), the mean error (ME), and the root mean square error (RMSE). The corresponding formula for each criterion is presented in Equations 4.4 to 4.6:

$$\mathrm{PC} = \frac{\sum \mathrm{C}}{\mathrm{N_S}}, \quad \text{for} \quad \mathrm{C} = \begin{cases} 1 & \text{if } \hat{\theta} = \theta \\ 0 & \text{if } \hat{\theta} \neq \theta \end{cases} \tag{4.4}$$

$$\mathrm{ME} = \frac{\sum(\hat{\theta} - \theta)}{N_S} \tag{4.5}$$

$$\mathrm{RMSE} = \sqrt{\frac{\sum(\hat{\theta} - \theta)^2}{N_S}} \tag{4.6}$$

where N_S is number of sample data matrices generated for each factor combination (100), $\hat{\theta}$ is the estimated number of factors, and θ is the population number of factors.

The PC criterion has boundaries of 0 and 1, with 0 indicating a total lack of accuracy and 1 reflecting perfect accuracy. In contrast, a 0 on the ME criterion shows a complete lack of bias, with negative and positive values indicating underfactoring and overfactoring, respectively. It is important to note that the ME cannot be used alone as a measure of method performance because errors of under- and overfactoring can compensate each other and give a false illusion of accuracy (this does not happen with the PC or RMSE criteria). In terms of the RMSE criterion, higher values signal larger deviations from the population number of factors, while a value of 0 indicates perfect accuracy. These three statistics were computed for each factor combination and were later averaged to obtain the values corresponding to each factor level.

All simulations were run under the MATLAB R2010a software (The MathWorks, Inc.). The polychoric correlations were computed according to the algorithms provided by Olsson (1979a). Also, the maximum number of factors possible (variables – 1) was extracted in order to compute the eigenvalues for PA-MRFA, as described in Timmerman and Lorenzo-Seva (2011).

4.6. Results

There were 172,431 non-Gramian polychoric matrices out of a total of 486,000 (35.5%). A multiple linear regression with the number of non-Gramian matrices per factor combination as the dependent variable and the 7 data factors as the independent variables showed that the sample size (β = -0.54) had the largest effect in the emergence of the non-Gramian polychoric matrices. In all, there were 114,940 non-Gramian matrices for N = 100 (66.6%), 50,154 for N = 300 (29.1%), and 7,337 for N = 1,000 (4.3%)[2].

A comparison of the two approaches used to work with the non-Gramian polychoric matrices yielded very similar results. Because MRFA requires Gramian correlation matrices the comparison could only be carried out for PCA extraction. The first approach of smoothing the non-Gramian matrices produced values of 0.56, -0.83, and 1.27, for the PC, ME, and RMSE performance criteria, respectively. Similarly, the second approach of using the positive and negative eigenvalues without any treatment yielded values of 0.56, -0.71 and 1.29, for the same three criteria. As can be seen by these results, both approaches produced nearly identical levels of accuracy. However, because PA-MRFA requires Gramian matrices, from this point forward in the manuscript all the results will be given for the smoothed non-Gramian matrices so that PA-PCA and PA-MRFA can be compared on the same input data.

An overall assessment of the performance of the PA methods is presented in Table 4.2. The first block of results in Table 4.2 shows that the polychoric correlations performed better than the Pearson correlations for both Gramian (e.g., PC[Paρ] = 0.74 > PC[PAr] = 0.66) and non-Gramian (e.g., PC[Paρ] = 0.56 > PC[PAr] = 0.46) polychoric matrices, leading to an

[2] The cases for the unidimensional portion of the design were not included in the computation of the global results presented in this section. This is because the unidimensional condition makes a not-completely crossed factorial design (there are no factor correlations), and thus can't be used in any of the inferential analysis (Multiple Linear Regressions or ANOVAs) or be distributed equally across all the cells of the design. Consequently, the decision was made to treat these results separately. The performance of PA in the unidimensional condition can be found on the 1-factor column of Tables 5 and 6.

overall better performance with polychoric correlations (e.g., PC[Paρ] = 0.68 > PC[PAr] = 0.59). Additionally, PCA extraction produced more accurate estimations than MRFA (e.g., RMSE[PA-PCA] = 0.85 < RMSE[PA-MRFA] = 1.18). A closer look reveals that PA-PCA and PA-MRFA performed very similarly for the cases with non-Gramian polychoric matrices (e.g., RMSE[PA-PCA] = 1.39 ≈ RMSE[PA-MRFA] = 1.42), and very differently for those with Gramian matrices (e.g., RMSE[PA-PCA] = 0.55 << RMSE[PA-MRFA] = 1.04). As will be seen later on, this is because PA-MRFA performed very closely to PA-PCA with N = 100, a condition that is overrepresented for non-Gramian matrices (66.6%), and substantially worse for N = 300 and N = 1,000, conditions which are overrepresented for Gramian matrices (85.0%). The final results for the first block of Table 4.2 show that the mean eigenvalue criteria performed better than the 95th percentile for both cases of Gramian (e.g., PC[PA_m] = 0.73 > PC[PA_{95}] = 0.67) and non-Gramian (e.g., PC[PA_m] = 0.52 > PC[PA_{95}] = 0.49) polychoric matrices, resulting in a better total performance (e.g., PC[PA_m] = 0.66 > PC[PA_{95}] = 0.61). In general, the 95th percentile tended to underextract more markedly than the mean criteria (ME[PA_{95}] = -0.91 < ME[PA_m] = -0.37).

The second block of results presented in Table 4.2 shows the performance for the 8 PA variants. Here it can be seen that PA-PCAρ_m was the most accurate method (e.g., RMSE[PA-PCAρ_m] = 0.70, lowest overall value) and PA-MRFAr_{95} the least accurate (e.g., RMSE[PA-MRFAr_{95}] = 1.50, highest overall value). In addition, all the PA methods showed a tendency to underextract (all Mes were negative) and the performance was always better for those cases with Gramian polychoric matrices. The overall poorer performance for all PA methods in the cases with non-Gramian polychoric matrices was expected because these matrices occur disproportionally at smaller sample sizes, where the estimations are generally less accurate. However, In order to determine if these differences were also due to a different behavior of PA with non-Gramian matrices, the performances of the 4 polychoric-based PA

methods were compared for those factor combinations where there were at least 10 Gramian and 10 non-Gramian replications. In total, 635 factor combinations met this criterion. For these 635 combinations, the Pearson correlation between the PC for the Gramian and non-Gramian matrices was computed. The mean correlation between the PC values for the 4 polychoric-based methods was 0.97, with a minimum of 0.96. In addition, the absolute mean difference between the total Gramian PC and the total non-Gramian PC for the 4 polychoric methods was 0.01, with a maximum of 0.02. These results indicate that the accuracy of PA is not affected in a noticeable way by the occurrence of non-Gramian polychoric matrices.

Table 4.2

Overall Parallel Analysis Performance

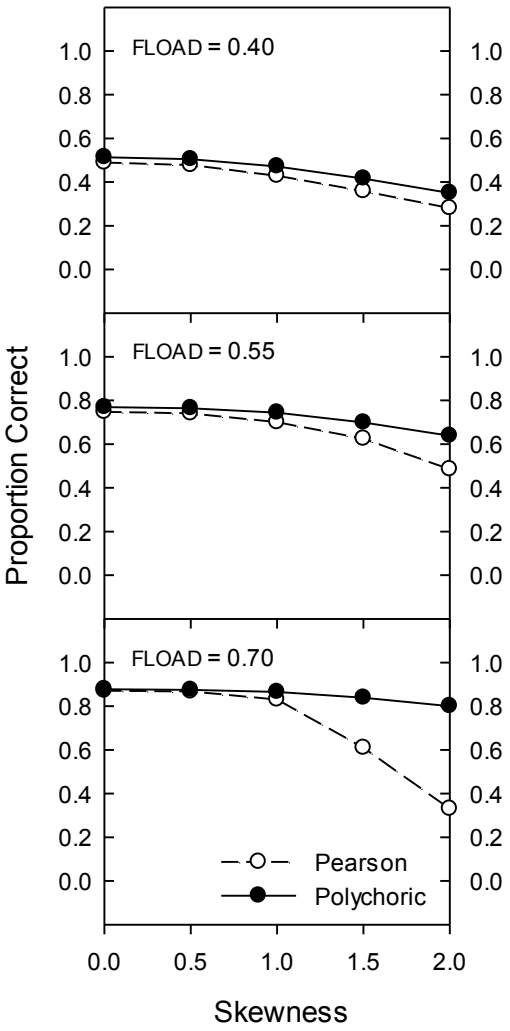
	Gramian PM (N = 313,569)			Non-Gramian PM (N = 172,431)			Total (N = 486,000)		
	PC	ME	RMSE	PC	ME	RMSE	PC	ME	RMSE
PAr	0.66	-0.59	0.90	0.46	-0.72	1.52	0.59	-0.64	1.12
Paρ	0.74	-0.51	0.70	0.56	-0.86	1.29	0.68	-0.63	0.91
PA-PCA	0.78	-0.27	0.55	0.51	-0.68	1.39	0.68	-0.41	0.85
PA-MRFA	0.62	-0.84	1.04	0.51	-0.91	1.42	0.58	-0.86	1.18
Pa_m	0.73	-0.35	0.68	0.52	-0.41	1.28	0.66	-0.37	0.90
PA_{95}	0.67	-0.76	0.91	0.49	-1.18	1.53	0.61	-0.91	1.13
$PA\text{-}PCAr_m$	0.76	-0.09	0.56	0.45	-0.11	1.42	0.65	-0.10	0.87
$PA\text{-}PCAr_{95}$	0.75	-0.39	0.65	0.45	-0.94	1.58	0.64	-0.58	0.98
$PA\text{-}PCA\rho_m$	0.82	-0.15	0.45	0.58	-0.48	1.15	0.73	-0.27	0.70
$PA\text{-}PCA\rho_{95}$	0.79	-0.43	0.55	0.54	-1.19	1.40	0.70	-0.70	0.86
$PA\text{-}MRFAr_m$	0.63	-0.63	0.99	0.48	-0.50	1.40	0.58	-0.58	1.13
$PA\text{-}MRFAr_{95}$	0.51	-1.27	1.40	0.44	-1.34	1.68	0.48	-1.29	1.50
$PA\text{-}MRFA\rho_m$	0.72	-0.51	0.73	0.58	-0.55	1.16	0.67	-0.52	0.88
$PA\text{-}MRFA\rho_{95}$	0.63	-0.94	1.05	0.53	-1.24	1.45	0.59	-1.05	1.19

Note. PM = polychoric matrix; PC = proportion correct; ME = mean error; RMSE = root mean square error; PA = parallel analysis; PCA = principal component analysis; MRFA = minimum rank factor analysis; r = Pearson correlation; ρ = polychoric correlation; m = mean eigenvalue; 95 = 95% eigenvalue.

As a means to summarize and better understand the results of the simulation study, a mixed factorial analysis of variance (ANOVA) was performed with the 3 method factors as the within- subject variables, the 7 data factors as the between-subject factors, and the proportion of correct estimates as the dependent variable. Due to the large sample size, most

of the effects were significant. For this reason, the partial eta squared (η_P^2) measure of effect size was chosen to establish the impact of the independent variables. According to Cohen (1988), values of 0.01 represent small effects, 0.06 medium effects, and 0.14 or more, large effects. Following this guide, the correlation type ($\eta_P^2 = 0.21$) and the extraction method ($\eta_P^2 = 0.23$) had large effect sizes, while the eigenvalue percentile ($\eta_P^2 = 0.06$) had a medium effect. In addition, a cutoff of ≥ 0.14 was used to establish the most salient interactions. In total, 3 within-subjects interactions reached this effect size: extraction method x variables per factor ($\eta_P^2 = 0.32$), correlation type x skewness ($\eta_P^2 = 0.20$) and correlation type x skewness x factor loading ($\eta_P^2 = 0.16$; see Figure 4.2). Two additional interactions, extraction method x number of variables per factor x sample size ($\eta_P^2 = 0.13$; see Figure 4.3) and extraction method x number of variables per factor x factor loading x factor correlation ($\eta_P^2 = 0.12$; see Figure 4.4), were included because they were theoretically and practically relevant and had an effect size near the cutoff of 0.14. Both double interactions (extraction method x number of variables per factor and correlation type x skewness) will be discussed in the context of the higher order interactions that include them.

The triple interaction of correlation type x skewness x factor loading presented in Figure 4.2 can be explained in two parts. First, the correlation type x skewness double interaction can be seen in each of the three blocks of Figure 4.2: as the skewness level increases the superiority of the polychoric methods over the Pearson methods becomes larger. Second, the correlation type x skewness interaction is also affected by the factor loadings: as the factor loadings increase the superiority of the polychoric methods with higher skewness becomes larger. In general, the Pearson and polychoric methods perform similarly for skewness levels of 0.00 to ±1.00, while the difference becomes more markedly with higher factor loadings and skewness levels of ±1.50, and especially, of ±2.00.



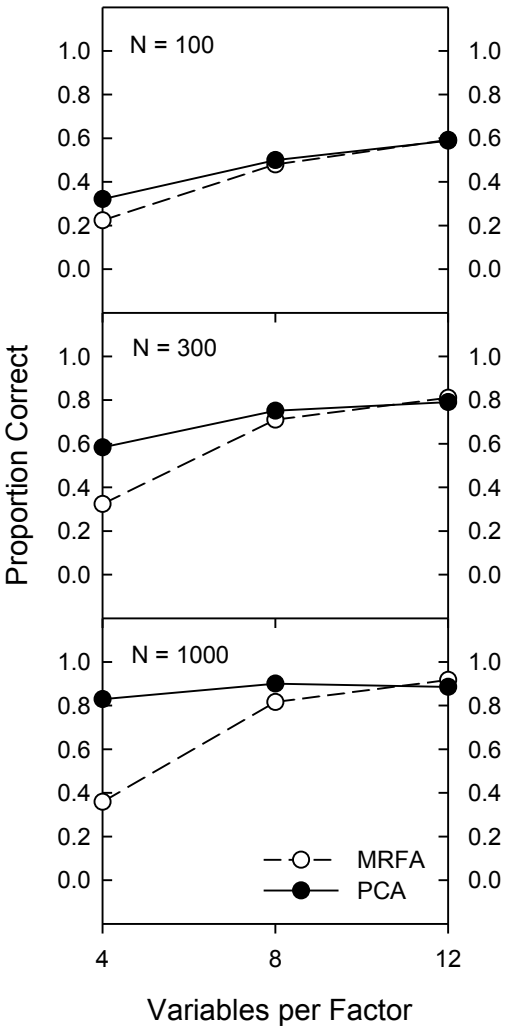
Note. FLOAD = factor loading

Figure 4.2: *Correlation Type x Skewness x Factor Loading Interaction*

A closer look at the performance of PAr reveals that the method is ineffective with large levels of skewness because of the emergence of difficulty factors. For example, with factor loadings of 0.70 and skewness of ±2.00, PAr overextracts in 44% of the cases, while Paρ only does so in 1% of the cases; this difference in overextractions explains almost completely the difference in proportion of correct estimates between the two methods (PC[Paρ] – PC[PAr] = 0.80 – 0.33 = 0.47). Moreover, the results from other less salient interactions (not mentioned here because of space constraints) indicate that with large levels of skewness, PAr has the undesirable property of becoming less accurate as the structures

become more robust or well-defined. For example, with |skewness| ≥ 1.50, factor loadings of 0.70, sample size of 100 and 8 variables per factor, PC(PAr) = 0.58 (15% overextractions), while PC(Paρ) = 0.68 (1% overextractions). As the sample size increases from 100 to 1,000, PC(PAr) actually decreases to 0.52 (45% overextractions) while PC(Paρ) increases to 1.00. With an additional increase to 12 variables per factor, PC(PAr) further decreases to 0.22 (78% overextractions), while PC(Paρ) remains at 1.00, indicating perfect accuracy.

The triple interaction of extraction method x variables per factor x sample size is presented next in Figure 4.3. This triple interaction includes the double interaction of extraction method x variables per factor, which can be seen in each of the 3 blocks included in the figure: PA-PCA is notably superior to PA-MRFA with 4 variables per factor, slightly/moderately superior with 8 variables per factor, and equally as accurate or slightly inferior with 12 variables per factor. In other words, there is a notable difference in accuracy between PA-PCA and PA-MRFA with 4 variables per factor, but this difference is reduced (and sometimes slightly reversed) as the number of variables per factor increases. The triple interaction is then produced by the interaction of the sample size with the other two independent variables: as the sample size increases, the superiority of PA-PCA grows substantially with 4 variables per factor, grows moderately with 8 variables per factor, and grows slightly in the opposite direction with 12 variables per factor. In general, the most prominent feature of this interaction is that with 4 variables per factor the accuracy of PA-MRFA doesn't improve nearly as much as does the accuracy of PA-PCA when the sample size increases.

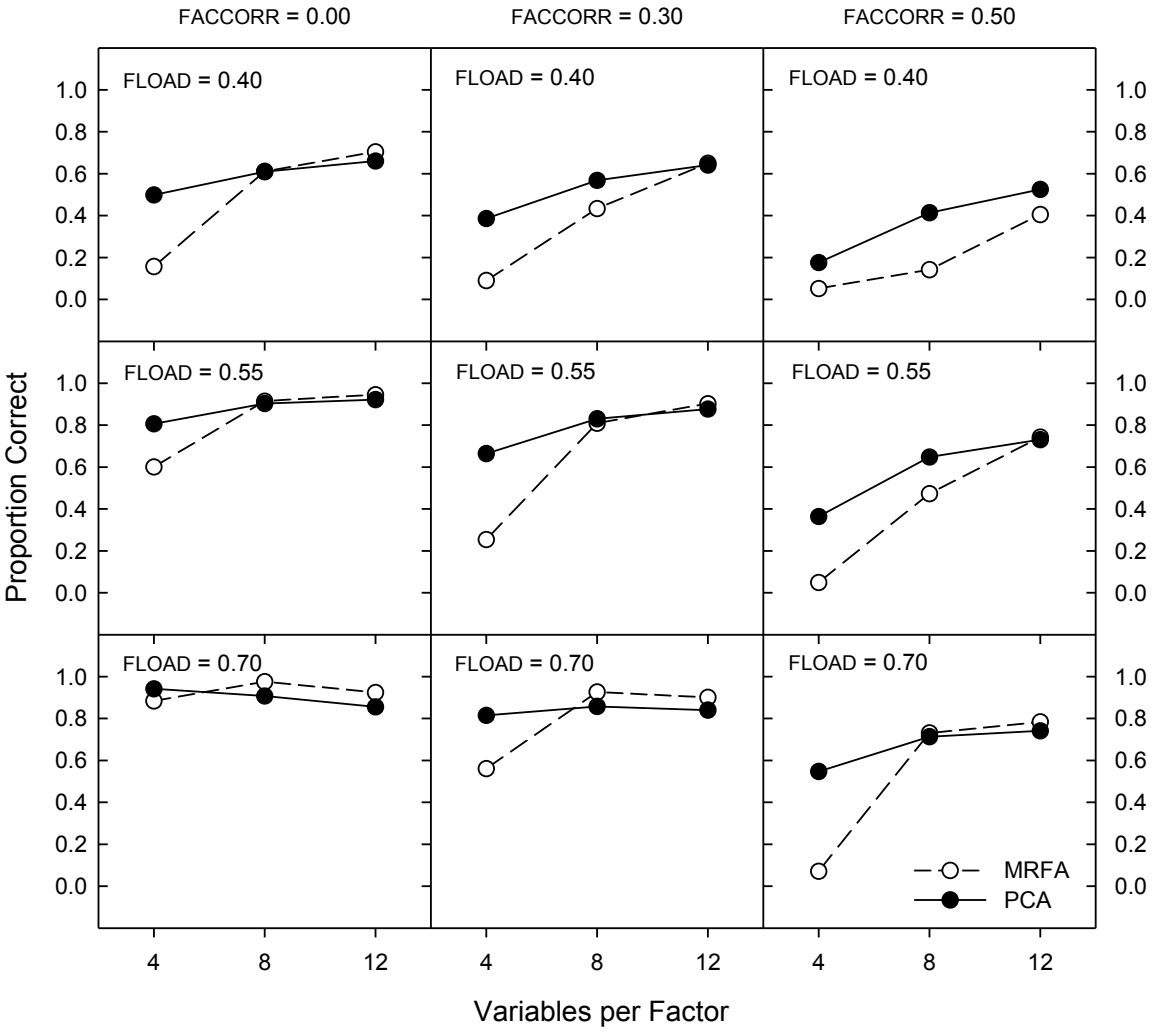


Note. N = sample size; MRFA = minimum rank factor analysis; PCA = principal component analysis

Figure 4.3: *Extraction Method x Variables per Factor x Sample Size Interaction*

The quadruple interaction of extraction method x number of variables per factor x factor loading x factor correlation is shown next in Figure 4.4. This interaction has two notable features. First, the extraction method x number of variables per factor double interaction can be seen in the majority of the 9 blocks: PA-MRFA is substantially inferior to PA-PCA with 4 variables per factor, but has much closer accuracy levels, and in some cases is even slightly superior, with 8 and 12 variables per factor. Second, PA-MRFA gradually closes the gap in accuracy with PA-PCA, and sometimes surpasses it, as the factor loadings increase and the factor correlations decrease (some exceptions to this trend occur for 4

variables per factor). This means that PA-MRFA is comparatively at its peak with factor loadings of 0.70 and factor correlations of 0.00.



Note. FLOAD = factor loading; FACCORR = factor correlation; MRFA = minimum rank factor analysis; PCA = principal component analysis

Figure 4.4: *Extraction Method x Number of Variables per Factor x Factor Loading x Factor Correlation Interaction*

The quadruple interaction presented in Figure 4.4 is especially important in order to put the findings of Timmerman and Lorenzo-Seva (2011) in the context of the results of this study. According to their results, PA-PCA and PA-MRFA performed equally well with 5 variables per factor, a finding that would seem to contradict those of the current study where

PA-MRFA performed substantially worse with a comparable small number (4) of variables per factor. However, a look at Figure 4.4 reveals that PA-PCA and PA-MRFA had very similar levels of accuracy with 4 variables per factor in the specific condition of high factor loadings (0.70) and zero factor correlations. Because Timmerman and Lorenzo-Seva did not study the performance of PA for correlated structures and because they kept the major factor loadings constant at 0.71, the findings of their research are actually in line with those of the current study for this particular condition. On the other hand, the results of this study show that for the other combinations of factor loading x factor correlation, PA-MRFA is substantially less accurate than PA-PCA with a small number of variables per factor.

Moving on to the between-subjects factors, separate ANOVAs (see Table 4.3) were performed for each PA method in order to determine the saliency of the different independent variables that were manipulated. The dependent variable in the ANOVAs was the proportion of correct estimates while the independent variables were the 7 data factors. According to the average effect sizes ($\bar{\eta}_P^2$), the most important variables were: the factor loading ($\bar{\eta}_P^2 = 0.21$), the number of variables per factor ($\bar{\eta}_P^2 = 0.20$), the sample size ($\bar{\eta}_P^2 = 0.19$), and the factor correlation ($\bar{\eta}_P^2 = 0.17$), all of which had average effect sizes of large magnitude. A second group of variables that had a medium impact on the performance of PA was composed of the number of factors ($\bar{\eta}_P^2 = 0.10$) and the skewness of the ordinal variables ($\bar{\eta}_P^2 = 0.09$). The last data factor, the number of response categories, only had a small impact ($\bar{\eta}_P^2 = 0.02$) on the accuracy of the PA methods. Of particular note was that the number of variables per factor had a much larger effect for PA-MRFA than for PA-PCA ($\bar{\eta}_P^2 = 0.33$ vs. 0.08), and that the skewness of the ordinal variables was more relevant for PA-PCA ($\bar{\eta}_P^2 = 0.19$ vs. 0.11).

Table 4.3

Univariate ANOVA Effect Sizes for the Parallel Analysis Methods

	Main Effects						
Method	N	FLOAD	VARFAC	FAC	FACCORR	RESCAT	SKEW
PA-PCAr_m	0.15	0.10	0.01	0.10	0.07	0.01	0.20
PA-PCAr_{95}	0.22	0.12	0.07	0.13	0.16	0.02	0.18
PA-PCAρ_m	0.26	0.24	0.06	0.12	0.08	0.02	0.03
PA-PCAρ_{95}	0.35	0.25	0.16	0.15	0.17	0.03	0.04
PA-MRFAr_m	0.08	0.17	0.22	0.09	0.16	0.02	0.11
PA-MRFAr_{95}	0.08	0.25	0.44	0.05	0.25	0.02	0.11
PA-MRFAρ_m	0.17	0.24	0.23	0.11	0.18	0.01	0.02
PA-MRFAρ_{95}	0.21	0.32	0.43	0.08	0.27	0.01	0.02
Average	0.19	0.21	0.20	0.10	0.17	0.02	0.09

Note. N = sample size; FLOAD =factor loading; VARFAC = variables per factor; FAC = number of factors; FACCORR = factor correlation; RESCAT = response categories; SKEW = skewness; PA = parallel analysis; PCA = principal component analysis; MRFA = minimum rank factor analysis; r = Pearson correlation; ρ = polychoric correlation; m = mean eigenvalue; 95 = 95% eigenvalue. Large effect sizes (≥ 0.14) are highlighted in grey shade.

The performance of the PA methods across the different levels of skewness is presented next in Table 4.4. In the case of PCA extraction, the performance of PA with Pearson and polychoric correlations was nearly identical for moderate levels (up to ±1.00) of skewness (e.g., with skewness of ±0.50: RMSE[PA-PCAr_{95}] = 0.68 ≈ RMSE[PA-PCAρ_{95}] = 0.70), while the polychoric methods were substantially superior for largely skewed data (e.g., with skewness of ±2.00: RMSE[PA-PCAρ_{95}] = 1.17 << RMSE[PA-PCAr_{95}] = 1.62). Regarding MRFA extraction, the polychoric methods were marginally superior to the Pearson methods with unskewed data (PC[PA-MRFAρ_{95}] = 0.63 > PC[PA-MRFAr_{95}] = 0.58), and the difference in accuracy grew gradually as the levels of skewness increased (e.g., with skewness of ±1.00: PC[PA-MRFAρ_{95}] = 0.61 > PC[PA-MRFAr_{95}] = 0.52). In general, the polychoric methods tended to underfactor more as the levels of skewness increased (ME[PA-PCAρ_{95}] = -0.55, -0.56, -0.63, -0.78 and -0.98 for skewness levels of 0.00, ±0.50, ±1.00, ±1.50 and ±2.00, respectively), while the Pearson methods had an irregular pattern due to the emergence of difficulty factors with large skewness (ME[PA-PCAr_{95}] = -0.53, -0.54, -0.64, -

0.67 and -0.55 for skewness levels of 0.00, ±0.50, ±1.00, ±1.50 and ±2.00, respectively). In general, PA-PCAρ_m was the most accurate method across the levels of skewness (e.g., for skewness of ±1.50: PC[PA-PCAρ_m] = 0.71, highest value for any method).

Table 4.4

Parallel Analysis Performance Across the Different Levels of Skewness

	Skewness				
Method	0.00	±0.50	±1.00	±1.50	±2.00
	Proportion Correct (PC)				
PA-PCAr_m	**0.79**	**0.78**	0.74	0.57	0.37
PA-PCAr_{95}	0.76	0.76	0.72	0.58	0.40
PA-PCAρ_m	**0.79**	**0.78**	**0.75**	**0.71**	**0.64**
PA-PCAρ_{95}	0.76	0.75	0.73	0.68	0.61
PA-MRFAr_m	0.68	0.67	0.63	0.54	0.37
PA-MRFAr_{95}	0.58	0.57	0.52	0.44	0.31
PA-MRFAρ_m	0.71	0.70	0.69	0.65	0.60
PA-MRFAρ_{95}	0.63	0.62	0.61	0.57	0.53
	Mean Error (ME)				
PA-PCAr_m	-0.21	-0.20	-0.23	-0.10	0.25
PA-PCAr_{95}	-0.53	-0.54	-0.64	-0.67	-0.55
PA-PCAρ_m	-0.22	-0.23	-0.25	-0.29	-0.36
PA-PCAρ_{95}	-0.55	-0.56	-0.63	-0.78	-0.98
PA-MRFAr_m	-0.55	-0.56	-0.62	-0.65	-0.55
PA-MRFAr_{95}	-1.09	-1.11	-1.25	-1.44	-1.59
PA-MRFAρ_m	-0.49	-0.50	-0.51	-0.54	-0.58
PA-MRFAρ_{95}	-0.94	-0.96	-0.99	-1.10	-1.24
	Root Mean Square Error (RMSE)				
PA-PCAr_m	**0.54**	**0.56**	0.67	1.02	1.55
PA-PCAr_{95}	0.66	0.68	0.80	1.13	1.62
PA-PCAρ_m	0.55	**0.56**	**0.64**	**0.78**	**0.96**
PA-PCAρ_{95}	0.68	0.70	0.79	0.95	1.17
PA-MRFAr_m	0.87	0.89	1.01	1.25	1.65
PA-MRFAr_{95}	1.22	1.24	1.39	1.64	1.99
PA-MRFAρ_m	0.77	0.79	0.84	0.94	1.09
PA-MRFAρ_{95}	1.06	1.09	1.14	1.25	1.41

Note. PA = parallel analysis; PCA = principal component analysis; MRFA = minimum rank factor analysis; r = Pearson correlation; ρ = polychoric correlation; m = mean eigenvalue; 95 = 95% eigenvalue; best column values are shown in boldface (highest PC and lowest RMSE); values similar to the best column values are highlighted in grey shade (within 0.05 of the highest PC and 0.10 of the lowest RMSE).

The performance of the PA methods at each level of the remaining independent variables is shown in Tables 5 and 6. The commentary of these results will be guided by the findings from the previous ANOVAs and of other PA research. Due to the relevancy of the skewness factor in the performance of PA with Pearson and polychoric correlations, the results have been divided in two sections: (a) for moderate levels of skewness (0.00 to ± 1.00; see Table 4.5), and (b) for largely skewed variables (± 1.50 to ± 2.00; see Table 4.6).

The most notable findings for the independent variables included in Tables 5 and 6 are discussed below in order. First, the PA methods tended to perform better in the expected conditions of larger sample size, higher factor loadings, more variables per factor, less number of factors, lower factor correlations and more response categories (e.g., for sample size of 1000 [Table 4.5]: PC[PA-PCAρ_m] = 0.96, highest for all sample size levels). There were some exceptions, however, with large skewness and Pearson correlations (e.g., for factor loadings of 0.40, 0.55 and 0.70 [Table 4.6]: PC[PA-PCAr_m] = 0.40, 0.61 and 0.41, respectively), which can be attributed to the overfactoring produced by the emergence of the difficulty factors (e.g., for factor loadings of 0.70 [Table 4.6]: ME[PA-PCAr_m] = +0.71).

Second, the mean eigenvalue criteria worked better for PCA extraction when the factors were correlated (e.g., for factor correlation of 0.50 [Table 4.5]: PC[PA-PCAρ_m] = 0.66 > PC[PA-PCAρ_{95}] = 0.58), while the 95th percentile lead to slightly more accurate estimations for orthogonal structures (e.g., for factor correlation of 0.00 [Table 4.5]: PC[PA-PCAρ_{95}] = 0.87 > PC[PA-PCAρ_m] = 0.85). In the case of MRFA extraction the mean eigenvalue criteria was almost uniformly the most accurate for any level of factor correlation (e.g., for factor correlation of 0.30 [Table 4.5]: PC[PA-MRFAρ_m] = 0.75 > PC[PA-MRFAρ_{95}] = 0.65).

Third, the maximum levels of accuracy were generally achieved in the unidimensional condition (e.g., for 1 factor [Table 4.6]: PC[PA-MRFAρ_m] = 0.92, highest PC value of this method for any factor level).

Fourth, the gains in accuracy for the number of response categories were maximal when going from 2 to 3 scale points (e.g., for 2 and 3 response categories [Table 4.6]: RMSE[PA-MRFAr_{95}] = 1.68 and 1.45, respectively; biggest reduction in RMSE for consecutive scale points), and they got gradually smaller as the number of response categories increased (e.g., for 4 and 5 response categories [Table 4.6]: RMSE[PA-MRFAr_m] = 1.37 and 1.31, respectively; smallest reduction in RMSE for consecutive scale points).

Fifth, all the PA methods tended to underestimate the number of factors (most Mes are negative), especially if the structures were not very robust or well-defined (e.g., for sample size of 100, factor loading of 0.40, 4 variables per factor and factor correlation of 0.50 [Table 4.5]: ME[PA-PCAρ_{95}] = -1.21, -1.06, -1.05 and -1.06, respectively; greatest underestimations for each independent variable).

Sixth, the most salient independent variables for each PA method (see Table 4.3) produced, as expected, the overall lowest and highest accuracy levels (e.g., for 4 and 12 variables per factor [Table 4.5]: PC[PA-MRFAr_{95}] = 0.19 and 0.82; lowest and highest PC for this PA method with moderate skewness).

Table 4.5

Parallel Analysis Performance Across the Different Levels of the Independent Variables with Skewness of 0.00 to ±1.00

	Sample Size			Factor Loading			Variables x Factor			Number of Factors				Factor Correlation			Response Categories			
Method	100	300	1000	0.40	0.55	0.70	4	8	12	1	2	4	6	0.00	0.30	0.50	2	3	4	5
										Proportion Correct (PC)										
PA-PCAr$_m$	**0.57**	0.79	0.95	0.56	**0.83**	**0.92**	**0.67**	**0.80**	0.84	0.96	**0.90**	**0.77**	**0.65**	0.85	**0.81**	**0.66**	**0.71**	**0.77**	**0.80**	**0.81**
PA-PCAr$_{95}$	0.51	0.78	0.94	0.55	0.79	0.90	0.59	0.79	**0.86**	**0.97**	0.88	0.73	0.62	**0.87**	0.78	0.58	0.69	0.74	0.77	0.78
PA-PCAρ$_m$	0.56	**0.80**	**0.96**	**0.57**	**0.83**	**0.92**	**0.67**	**0.80**	0.85	0.96	**0.90**	**0.77**	**0.65**	0.85	**0.81**	**0.66**	**0.71**	**0.77**	**0.80**	**0.81**
PA-PCAρ$_{95}$	0.50	0.79	0.95	0.55	0.79	0.90	0.59	0.79	**0.86**	**0.97**	0.88	0.73	0.62	**0.87**	0.78	0.58	0.68	0.74	0.77	0.79
PA-MRFAr$_m$	0.53	0.69	0.77	0.43	0.72	0.84	0.40	0.74	0.85	0.91	0.78	0.65	0.56	0.81	0.71	0.47	0.60	0.66	0.69	0.70
PA-MRFAr$_{95}$	0.43	0.60	0.64	0.32	0.58	0.76	0.19	0.65	0.82	0.76	0.63	0.55	0.49	0.75	0.57	0.35	0.48	0.55	0.59	0.60
PA-MRFAρ$_m$	0.54	0.73	0.82	0.48	0.76	0.86	0.46	0.78	**0.86**	0.95	0.82	0.69	0.59	0.84	0.75	0.51	0.65	0.70	0.72	0.73
PA-MRFAρ$_{95}$	0.46	0.66	0.74	0.39	0.66	0.81	0.29	0.72	0.85	0.84	0.71	0.61	0.54	0.80	0.65	0.41	0.58	0.62	0.64	0.65
										Mean Error (ME)										
PA-PCAr$_m$	-0.45	-0.13	-0.05	-0.27	-0.24	-0.12	-0.55	-0.10	0.02	0.04	0.04	-0.15	-0.52	0.11	-0.11	-0.64	-0.21	-0.21	-0.21	-0.20
PA-PCAr$_{95}$	-1.17	-0.43	-0.10	-1.06	-0.46	-0.19	-1.06	-0.42	-0.22	-0.02	-0.12	-0.51	-1.08	-0.22	-0.45	-1.04	-0.71	-0.57	-0.51	-0.48
PA-PCAρ$_m$	-0.51	-0.14	-0.05	-0.30	-0.26	-0.14	-0.55	-0.13	-0.02	0.04	0.03	-0.17	-0.56	0.10	-0.12	-0.67	-0.28	-0.23	-0.22	-0.20
PA-PCAρ$_{95}$	-1.21	-0.43	-0.11	-1.06	-0.47	-0.21	-1.05	-0.44	-0.24	-0.02	-0.12	-0.52	-1.11	-0.22	-0.46	-1.06	-0.75	-0.58	-0.51	-0.47
PA-MRFAr$_m$	-0.69	-0.53	-0.51	-0.80	-0.56	-0.37	-1.37	-0.30	-0.05	0.02	-0.11	-0.54	-1.08	-0.05	-0.41	-1.26	-0.63	-0.58	-0.56	-0.54
PA-MRFAr$_{95}$	-1.42	-1.01	-1.01	-1.86	-1.02	-0.56	-2.37	-0.76	-0.31	-0.24	-0.49	-1.12	-1.83	-0.64	-1.06	-1.74	-1.36	-1.16	-1.06	-1.01
PA-MRFAρ$_m$	-0.65	-0.44	-0.42	-0.70	-0.48	-0.32	-1.21	-0.24	-0.05	0.01	-0.09	-0.45	-0.95	0.00	-0.33	-1.17	-0.52	-0.50	-0.50	-0.48
PA-MRFAρ$_{95}$	-1.34	-0.82	-0.72	-1.64	-0.81	-0.44	-2.01	-0.61	-0.27	-0.15	-0.36	-0.92	-1.61	-0.46	-0.83	-1.60	-1.07	-0.97	-0.92	-0.89
										Root Mean Square Error (RMSE)										
PA-PCAr$_m$	**1.12**	0.51	0.13	1.14	**0.43**	**0.19**	**0.87**	**0.50**	0.40	0.13	0.24	0.54	**0.99**	0.41	**0.46**	**0.89**	**0.76**	**0.59**	0.52	0.49
PA-PCAr$_{95}$	1.40	0.60	0.16	1.32	0.57	0.26	1.21	0.57	0.37	**0.06**	**0.21**	0.66	1.28	**0.38**	0.59	1.18	0.90	0.72	0.64	0.61
PA-PCAρ$_m$	1.13	**0.50**	**0.12**	**1.13**	**0.43**	**0.19**	**0.87**	**0.50**	0.38	0.13	0.23	**0.52**	**0.99**	0.39	**0.46**	0.90	**0.76**	**0.59**	**0.51**	**0.47**
PA-PCAρ$_{95}$	1.42	0.59	0.15	1.32	0.57	0.27	1.21	0.58	0.37	**0.06**	**0.21**	0.66	1.29	**0.38**	0.59	1.19	0.93	0.72	0.64	0.59
PA-MRFAr$_m$	1.24	0.83	0.69	1.57	0.76	0.44	1.71	0.66	0.39	0.21	0.40	0.88	1.48	0.54	0.76	1.47	1.11	0.93	0.85	0.80
PA-MRFAr$_{95}$	1.61	1.14	1.10	2.05	1.16	0.65	2.51	0.91	0.43	0.32	0.60	1.26	2.00	0.80	1.21	1.85	1.52	1.30	1.19	1.13
PA-MRFAρ$_m$	1.18	0.71	0.51	1.37	0.65	0.38	1.49	0.55	**0.36**	0.14	0.32	0.76	1.32	0.42	0.62	1.36	0.92	0.80	0.75	0.72
PA-MRFAρ$_{95}$	1.54	0.96	0.79	1.84	0.94	0.51	2.16	0.74	0.39	0.22	0.46	1.06	1.77	0.62	0.97	1.70	1.22	1.10	1.05	1.02

Note. PA = parallel analysis; PCA = principal component analysis; MRFA = minimum rank factor analysis; r = Pearson correlation; ρ = polychoric correlation; m = mean eigenvalue; 95 = 95% eigenvalue; best column values are shown in boldface (highest PC and lowest RMSE); values similar to the best column values are highlighted in grey shade (within 0.05 of the highest PC and 0.10 of the lowest RMSE); the 1-factor results are not averaged across the other variables.

Table 4.6

Parallel Analysis Performance Across the Different Levels of the Independent Variables with Skewness of ±1.50 to ±2.00

	Sample Size			Factor Loading			Variables x Factor			Number of Factors				Factor Correlation			Response Categories			
Method	100	300	1000	0.40	0.55	0.70	4	8	12	1	2	4	6	0.00	0.30	0.50	2	3	4	5
Proportion Correct (PC)																				
PA-PCAr$_m$	0.35	0.47	0.60	0.40	0.61	0.41	0.51	0.49	0.42	0.74	0.63	0.45	0.34	0.56	0.51	0.34	0.41	0.47	0.50	0.52
PA-PCAr$_{95}$	0.33	0.51	0.64	0.40	0.62	0.45	0.44	0.53	0.51	0.81	0.66	0.46	0.35	0.64	0.53	0.30	0.43	0.49	0.51	0.53
PA-PCAρ$_m$	**0.44**	**0.69**	**0.90**	**0.42**	**0.73**	**0.87**	**0.57**	**0.70**	0.75	0.91	**0.82**	**0.66**	**0.53**	0.75	**0.71**	**0.56**	**0.59**	**0.67**	**0.71**	**0.73**
PA-PCAρ$_{95}$	0.36	0.67	**0.90**	0.41	0.69	0.83	0.48	0.68	**0.77**	**0.94**	0.80	0.62	0.51	**0.78**	0.68	0.48	0.56	0.64	0.68	0.70
PA-MRFAr$_m$	0.34	0.47	0.55	0.29	0.55	0.52	0.26	0.55	0.55	0.80	0.61	0.43	0.33	0.61	0.48	0.27	0.39	0.45	0.48	0.50
PA-MRFAr$_{95}$	0.28	0.40	0.45	0.19	0.44	0.49	0.07	0.46	0.59	0.63	0.47	0.36	0.29	0.56	0.40	0.17	0.32	0.38	0.39	0.42
PA-MRFAρ$_m$	0.43	0.65	0.80	0.39	0.68	0.82	0.42	0.69	**0.77**	0.92	0.77	0.61	0.50	0.76	0.68	0.45	0.56	0.63	0.65	0.67
PA-MRFAρ$_{95}$	0.34	0.59	0.72	0.31	0.59	0.76	0.25	0.64	0.76	0.82	0.66	0.54	0.45	0.73	0.57	0.35	0.49	0.55	0.57	0.59
Mean Error (ME)																				
PA-PCAr$_m$	-0.47	0.17	0.52	-0.30	-0.18	0.71	-0.73	0.22	0.74	0.26	0.33	0.17	-0.27	0.56	0.20	-0.53	0.05	0.08	0.08	0.08
PA-PCAr$_{95}$	-1.61	-0.50	0.28	-1.50	-0.70	0.36	-1.53	-0.44	0.14	0.08	-0.03	-0.54	-1.27	-0.17	-0.48	-1.18	-0.81	-0.60	-0.54	-0.49
PA-PCAρ$_m$	-0.76	-0.16	-0.05	-0.37	-0.36	-0.24	-0.72	-0.22	-0.04	0.09	0.06	-0.24	-0.80	0.13	-0.21	-0.89	-0.44	-0.31	-0.28	-0.27
PA-PCAρ$_{95}$	-1.75	-0.70	-0.19	-1.49	-0.76	-0.39	-1.45	-0.73	-0.45	-0.04	-0.21	-0.81	-1.62	-0.46	-0.76	-1.42	-1.14	-0.87	-0.77	-0.73
PA-MRFAr$_m$	-0.77	-0.52	-0.51	-0.88	-0.71	-0.20	-1.69	-0.35	0.25	0.07	-0.07	-0.55	-1.17	0.03	-0.50	-1.32	-0.66	-0.59	-0.58	-0.56
PA-MRFAr$_{95}$	-1.88	-1.34	-1.33	-2.34	-1.47	-0.73	-2.94	-1.21	-0.39	-0.35	-0.69	-1.48	-2.37	-1.05	-1.45	-2.04	-1.77	-1.49	-1.43	-1.37
PA-MRFAρ$_m$	-0.83	-0.42	-0.43	-0.74	-0.55	-0.39	-1.27	-0.33	-0.08	0.02	-0.07	-0.49	-1.11	0.01	-0.39	-1.29	-0.64	-0.55	-0.52	-0.52
PA-MRFAρ$_{95}$	-1.80	-0.97	-0.74	-1.92	-1.02	-0.57	-2.18	-0.86	-0.47	-0.17	-0.42	-1.12	-1.97	-0.65	-1.04	-1.82	-1.37	-1.16	-1.10	-1.06
Root Mean Square Error (RMSE)																				
PA-PCAr$_m$	1.67	1.24	0.95	1.65	0.91	1.29	1.29	1.18	1.39	0.46	0.73	1.25	1.88	1.18	1.17	1.51	1.49	1.29	1.21	1.15
PA-PCAr$_{95}$	1.97	1.24	0.90	1.84	1.03	1.24	1.74	1.18	1.19	0.29	0.60	1.33	2.19	1.06	1.26	1.80	1.58	1.38	1.30	1.23
PA-PCAρ$_m$	**1.53**	**0.81**	**0.26**	**1.57**	**0.70**	**0.35**	**1.16**	**0.80**	0.65	0.23	0.38	**0.80**	**1.43**	0.68	**0.72**	**1.20**	**1.12**	**0.87**	**0.77**	**0.72**
PA-PCAρ$_{95}$	1.96	0.93	0.29	1.80	0.91	0.47	1.62	0.92	0.64	**0.12**	**0.35**	0.99	1.84	0.70	0.93	1.56	1.35	1.06	0.94	0.89
PA-MRFAr$_m$	1.74	1.39	1.23	2.12	1.18	1.06	2.16	1.15	1.05	0.44	0.74	1.42	2.20	1.14	1.34	1.88	1.68	1.45	1.37	1.31
PA-MRFAr$_{95}$	2.16	1.68	1.61	2.54	1.65	1.25	3.06	1.44	0.94	0.45	0.87	1.79	2.79	1.43	1.73	2.29	2.06	1.80	1.73	1.67
PA-MRFAρ$_m$	1.55	0.92	0.57	1.70	0.85	0.49	1.61	0.81	**0.62**	0.20	0.43	0.96	1.65	**0.67**	0.83	1.54	1.21	1.01	0.94	0.89
PA-MRFAρ$_{95}$	2.01	1.16	0.83	2.15	1.18	0.67	2.32	1.03	0.65	0.25	0.55	1.28	2.17	0.87	1.21	1.92	1.55	1.32	1.25	1.21

Note. PA = parallel analysis; PCA = principal component analysis; MRFA = minimum rank factor analysis; r = Pearson correlation; ρ = polychoric correlation; m = mean eigenvalue; 95 = 95% eigenvalue; best column values are shown in boldface (highest PC and lowest RMSE); values similar to the best column values are highlighted in grey shade (within 0.05 of the highest PC and 0.10 of the lowest RMSE); the 1-factor results are not averaged across the other variables.

4.7. Discussion

Horn's parallel analysis (PA) is currently one of the most accurate and recommended methods to assess data dimensionality (Hayton et al., 2004; Velicer et al., 2000; Zwick & Velicer, 1986), a critical phase of an exploratory factor analysis (Henson & Roberts, 2006). In recent years, the study of PA has extended to the determination of the number of factors with ordinal variables, typically encountered in the educational and psychological fields. Unfortunately, results from these studies with ordinal variables have produced unexpected findings as PA with Pearson correlations has performed as well or better than PA with the more theoretically appropriate polychoric correlations (Cho et al., 2009; Weng & Cheng, 2005). In the present study we identify several reasons for these unexpected results and conduct a comprehensive simulation study to evaluate more accurately the performance of PA with Pearson and polychoric correlations.

Regarding the main goal of this study, the comparison of PA with Pearson and polychoric correlations, the findings were two-fold: (1) PA with polychoric correlations performed similarly to PA with Pearson correlations for moderate levels of skewness (0.00 to ±1.00), thus extending the results of Cho et al. (2009) and Timmerman and Lorenzo-Seva (2011) for unskewed data; and (2) PA with polychorics was substantially more accurate for highly skewed ordinal variables (±1.50 to ±2.00) that had medium (0.55), and, especially, high (0.70) factor loadings, a novel finding of the current study. In addition, PA with Pearson correlations had the undesirable property of losing accuracy as the structures of highly skewed variables became more robust or well-defined (higher factor loadings, larger sample sizes and more variables per factor), while PA with polychorics was increasingly more accurate in these conditions. Overall, the results from a mixed ANOVA showed that the *type of correlation matrix* factor (Pearson vs. polychoric) had a large effect ($\eta_P^2 = 0.21$) in the accuracy of PA.

A key decision made in this study that enabled the emergence of these theoretically expected (Gorsuch, 1983; Olsson, 1979b), but previously unattained results, was the determination to analyze all polychoric correlation matrices, Gramian and non-Gramian. The non-Gramian polychoric matrices were analyzed by using a straightforward smoothing algorithm that eliminated all the negative eigenvalues and guaranteed that the PA rationale could be maintained exactly (as the eigenvalues were again related to the variance explained by the factor). The empirical results showed that this approach produced good PA estimations that were more accurate than those obtained with the originally Gramian Pearson matrices. Additionally, a second approach was also tested where the negative eigenvalues were not given any treatment. Although this latter method is based on a more liberal interpretation of Horn's PA, the empirical results showed that its performance was virtually identical to the more theoretically appropriate smoothing approach.

A secondary goal of this study was to determine the impact of the different independent variables, and their interactions, in the accuracy of the PA procedure. Concerning the *extraction method* within-subjects "method" factor, the results showed that PA with Principal Component Analysis (PCA) was more accurate than PA with Minimum Rank Factor Analysis (MRFA), and that the difference in performance could be categorized as large (η_P^2 = 0.23). PA with MRFA extraction tended to perform well with medium (8) and large (12) numbers of variables per factor and with orthogonal structures of highly loading (0.70) variables, but was generally ineffective with a small number (4) of variables per factor. The ANOVA results showed that the extraction method x number of variables per factor interaction had a large effect size (η_P^2 = 0.32), mostly because of the superiority of PCA extraction with a small number variables per factor. These results clarify and extend those of the Timmerman and Lorenzo-Seva (2011) study, where PCA and MRFA had similar levels of accuracy for uncorrelated structures of highly loading (0.71) variables. In terms of the final

method factor, the *eigenvalue percentile*, the results indicated that the mean of the random eigenvalues produced generally more accurate estimations than the 95th percentile, especially with correlated structures, as in Cho et al. (2009) and Crawford et al. (2010). According to the mixed ANOVA, the eigenvalue percentile had a medium effect in the performance of PA ($\eta_P^2 = 0.06$).

Regarding the between-subject factors, separate ANOVAs were performed to evaluate the saliency of the independent variables on the accuracy of the 8 variants of the PA method that were produced by the 2 x 2 x 2 (type of correlation matrix x extraction method x eigenvalue percentile) within-subjects design. According to the average effect sizes ($\bar{\eta}_P^2$) for the 8 methods, the most salient variables were: the *factor loading* ($\bar{\eta}_P^2 = 0.21$), the *number of variables per factor* ($\bar{\eta}_P^2 = 0.20$), the *sample size* ($\bar{\eta}_P^2 = 0.19$), and the *factor correlation* ($\bar{\eta}_P^2 = 0.17$), all of which had an average effect of large magnitude (≥ 0.14). Next in line were the *number of factors* ($\bar{\eta}_P^2 = 0.10$) and the *skewness* ($\bar{\eta}_P^2 = 0.09$), which had a medium effect (≥ 0.06), while the *number of response categories* only had a small impact ($\bar{\eta}_P^2 = 0.02$) on the performance of the PA methods. These results are generally in line with previous PA research with continuous and ordinal variables (e.g., Beauducel, 2001; Cho et al., 2009; Zwick & Velicer, 1986). It is worth noting that the number of variables per factor was especially relevant for PA with MRFA extraction, while the skewness was a more salient variable for PA with PCA extraction and Pearson correlations. Overall, the most accurate PA estimations were obtained with the combination of polychoric correlations, PCA extraction, and the mean of the random eigenvalues.

Although PCA is not an appropriate extraction method to estimate and interpret the factor structure of a set of variables, it works very well with PA to determine the number of common factors present in the data. In contrast to other PCA based methods such as Velicer's

minimum average partial (Garrido et al., 2011), PA with PCA does not appear to be biased in the conditions of low population loadings and/or a small number of variables per factor, where PCA is known to overestimate the variable saturation. The main reason that this variant of PA works so well is because whatever bias is introduced in the loadings by using PCA extraction it works similarly for the real and random data, so that when they are compared the bias is mostly eliminated or greatly reduced. It is this particular property of PA that also makes it perform well with Pearson correlations when the ordinal variables are not heavily skewed. In this case, the underestimation of the correlations and the loadings are, again, similar for both types of datasets, so that no discernible bias is introduced when the eigenvalues are compared. This does not happen, on the contrary, with Pearson correlations and heavily skewed ordinal variables. The difference in this latter case is that high loadings appear to be necessary for the difficulty factors to emerge (see Olsson, 1979b), and those will usually *only* occur for the real datasets, therefore introducing a bias in the dimensionality assessment when the eigenvalues from the real and random data are compared.

A final analysis of the performance of PA was carried out in order to determine the most effective methods according to the types of data that may be encountered in practice. Of the 7 between-subjects "data" factors that were manipulated in the current study, 4 can be considered structure factors (factor loadings, number of variables per factor, factor correlations and number of factors) and 3 sample factors (sample size, number of response categories and skewness). The levels of the structure factors are, of course, unknown to the researcher, while the sample levels are completely known after the data has been collected and may be used to determine the most appropriate PA variant for a particular dataset. Therefore, the performance of the PA methods was analyzed for each of the 60 combinations (3 x 4 x 5) of sample size x number of response categories x skewness. The results from this analysis indicated that the overall most effective PA variant (polychoric correlations + PCA

extraction + mean eigenvalue criterion) had a proportion of correct estimates that was never more than 0.01 below to that of the most accurate method for each factor combination. Similarly, the RMSE of this PA variant was never more than 0.03 above than that of the best performing method for each combination. In general, these results suggest that the overall most effective PA method may be used without a noticeable loss in accuracy for any of the sample characteristics that were investigated in this study.

There are some limitations in this study that should be noted. First, all the models had perfect simple structure with equal factor loadings, number of variables per factor, and factor correlations within cases. This strategy is usually preferred for simulation studies because it allows for the generation of data that have perfectly known dimensionalities in the population. However, these are also idealized models that are not likely to be encountered in practical settings. For this reason, the results from this study should be seen as a best-case scenario. Second, PA was the only factor retention method investigated. This decision was made due to the importance and complexity of the PA method and because the results from previous studies with ordinal variables had been equivocal and unexpected. However, this procedure should be tested alongside other factor retention techniques in the future. Last, the same set of population thresholds that were employed to categorize the "real" normal variables were used to obtain the "random" ordinal data. This population threshold procedure was used in order to reduce the simulation time considerably by not having to compute the criterion eigenvalues for each dataset. In practice, where the population thresholds are unknown, researchers would perform random column permutations of the real data matrix in order to obtain random variables with the same levels of skewness as those from the real dataset. As a means to determine the effect that the population threshold approach may have had on the results of this study, one full replication (5,400 factor combinations) was simulated using both procedures. For each of the 8 PA methods the Pearson correlation was

computed between the numbers of factors estimated by the random permutation and population threshold procedures. The mean of these 8 correlation coefficients was 0.99, with a minimum coefficient of 0.98. In addition, the mean absolute difference in the proportion of correct estimates across the two procedures was 0.002, with a maximum of 0.004. These results indicate that both data generating approaches lead to nearly identical estimations and accuracy levels, and therefore may be used indistinctively in simulation studies without a meaningful loss in generalizability.

Taking into consideration the combined results of the simulation study we propose the following guidelines to researchers who wish to use PA to determine the dimensionality of ordinal variables:

1. The method of choice for all types of data is PA with polychoric correlations, PCA extraction, and the mean eigenvalue criterion.
2. If PA with polychoric correlations is not available, PA with Pearson correlations, PCA extraction, and the mean eigenvalue criterion may be used for moderately skewed data (0.00 to ±1.00) without any loss in accuracy.
3. The non-Gramian polychoric matrices may be smoothed using the Eigenvalue method described in the Method section or can be factorized as they are without any transformation.
4. Random column permutations of the real data matrix are recommended in order to generate the random criterion variables in practice.

A final note of clarification on the use of PCA extraction and Pearson correlations with ordinal variables seems warranted. PA with PCA extraction performed relatively well across the numerous conditions that were evaluated in this study, while PA with Pearson correlations produced accurate dimensionality estimates as long as the variables were not

greatly skewed. These results do not imply, however, that the subsequent factor analysis following the dimensionality assessment phase may be performed with this type of extraction method and correlation matrix as well. As was argued earlier, the reason that PA works well in these cases is that the biases of PCA extraction and Pearson correlations will tend to affect the size of the real and random eigenvalues similarly, resulting in a "cancelling of errors" once these eigenvalues are *compared* to each other. However, if this extraction method and/or correlation matrix were used to estimate and interpret an *isolated* factor solution, the results would be biased and misleading. In particular, PCA extraction would strongly overestimate the factor loadings with a small number of variables per factor and/or low population loadings (Fabrigar et al., 1999; Widaman, 1993), while the Pearson correlations would produce downwardly biased factor loadings, especially with ordinal variables that had a small number of response categories (Babakus, et al., 1987; Bollen & Barb, 1981). So, even if PA with PCA extraction and Pearson correlations produced a correct dimensionality estimate, the factor-loading matrix would not be amenable to interpretation if it were estimated in the same manner. For this reason, once the number of factors has been determined, the actual factor-loading matrix to be interpreted should always be estimated, if possible, using polychoric correlations and a common factor extraction method such as unweighted least squares (ULS) or diagonally weighted least squares (DWLS).

4.8. Appendix

Table 4.7

Thresholds used to obtain the ordinal variables

	Thresholds			
RC	1	2	3	4
Skewness = 0.00				
2	0.0000			
3	-1.0000	1.0000		
4	-1.5000	0.0000	1.5000	
5	-1.8000	-0.6000	0.6000	1.8000
Skewness = 0.50				
2	0.3088			
3	-0.0236	0.7256		
4	-0.2057	0.3706	0.9809	
5	-0.3414	0.1642	0.6257	1.1645
Skewness = 1.00				
2	0.5936			
3	0.3195	0.9921		
4	0.1678	0.6873	1.2513	
5	0.0502	0.5117	0.9432	1.4462
Skewness = 1.50				
2	0.8416			
3	0.6131	1.1969		
4	0.4945	0.9299	1.4359	
5	0.4071	0.7827	1.1596	1.6186
Skewness = 2.00				
2	1.0518			
3	0.8518	1.3754		
4	0.7515	1.1341	1.5980	
5	0.6792	1.0043	1.3441	1.7703

Note. RC = response categories; negative skewness is obtained by changing the sign of the thresholds.

5. STUDY 3: USING FIT INDICES TO DETERMINE THE NUMBER OF FACTORS WITH ORDINAL VARIABLES

An early step in the process of construct validation consists in establishing the fit of an unrestricted "exploratory" factorial model for a prespecified number of common factors. Because this unrestricted model must show an acceptable fit in order to proceed with more stringent model testing, such as the fitting of a restricted "confirmatory" measurement model or a full structural equation "SEM" model, researchers have often recommended and used fit indices to determine the number of factors for the initial unrestricted model. Despite the logical appeal of this approach, little is known about the actual efficacy of fit indices in establishing data dimensionality. The present study aims to bridge this gap by systematically evaluating the accuracy of 4 commonly used fit indices – CFI, TLI, RMSEA, and SRMR – in determining the dimensionality of ordinal-level data across a wide range of conditions and cutoff values. According to the results of the simulation study, the CFI and TLI indices perform adequately at cutoff values of 0.95, while the RMSEA and SRMR indices are too erratic to be recommended for practical use. A two-method approach using the CFI/TLI indices and Horn's Parallel Analysis is proposed.

Researchers in the social sciences are often interested in the study of latent – unobserved – variables that are the causes of the behaviors they observe (Bollen, 2002). Latent variables such as aptitudes, feelings, and motives, in the context of a well-reasoned theory, have the potential to explain a wide array of behavioral processes using a relatively small number of constructs (Hoyle & Duvall, 2004). These latent variables, or factors, are random variables whose properties must be inferred indirectly using a statistical model that connects the latent variables to manifest – observed – variables, believed to be caused, at

least in part, by one or more factors (Mulaik & Millsap, 2000). The primary statistical tool for drawing such inferences is factor analysis, a technique that aims to describe the associations among a potentially large number of observed variables using a relatively small number of factors, the latent variables in the statistical model that represent the underlying processes that have been operating (Browne & Cudeck, 1992; Hoyle & Duvall, 2004).

A fundamental concern in factor analysis, and in extension, structural equation modeling (SEM), is the determination of the number of factors to retain for a group of variables of interest (Fabrigar, Wegener, MacCallum, & Strahan, 1999; Hayduk & Glaser, 2000; Hayton, Allen, & Scarpello, 2004; Henson & Roberts, 2006; Mulaik & Millsap, 2000; Schmitt, 2011). This decision is especially important because errors of underfactoring (extracting too few factors) or overfactoring (extracting too many factors) are likely to result in non-interpretable or unreliable factors (Fava & Velicer, 1992, 1996; Lee & Comrey, 1979; Wood, Tataryn, & Gorsuch, 1996), and can potentially mislead theory development efforts (Fabrigar et al., 1999). However, despite the substantial attention that this topic has received for more than half a century (e.g., Cattell, 1966; Horn, 1965; Kaiser, 1960; Steiger & Lind, 1980; Timmerman & Lorenzo-Seva, 2011; Velicer, 1976; Velicer, Eaton, & Fava, 2000), it continues to be a subject of considerable debate among quantitative methodologists (e.g., Bollen, 2000; Hayduk & Glaser, 2000; Herting & Costner, 2000; Mulaik & Millsap, 2000) and remains poorly understood by applied researchers (Fabrigar et al., 1999; Hayton et al., 2004; Henson & Roberts, 2006).

One of the important themes in the debate about the determination of the number of factors is the worrying disconnect that exists between the initial dimensionality assessment phase, usually performed with principal component analysis' (PCA) or traditional exploratory factor analysis' (EFA) techniques such the eigenvalue-greater-than-1 rule (Kaiser, 1960) or the scree test (Cattell, 1966), and the more stringent model validation phase, carried out under

the framework of confirmatory factor analysis (CFA) or full-blown SEM (Bollen, 2000). A typical scenario of well-planned research involves the determination of a clear and replicable EFA solution through a series of preliminary studies, followed by the fitting in a new sample of a CFA model based on the EFA solution (Ferrando & Lorenzo-Seva, 2000). What usually happens in this scenario is that the apparently robust EFA structure fits very badly when modeled through CFA (Ferrando & Lorenzo-Seva, 2000; Floyd & Widaman, 1995), a situation that could be due, at least in part, to an erroneous specification of the number of factors (Mulaik & Millsap, 2000). The researcher is then faced with the option of adding model parameters on an ad-hoc basis until the model shows an acceptable fit, a procedure that is not grounded in theory and will often capitalize on chance, or, on the other hand, he/she must drop the use of CFA and SEM altogether, an equally undesirable proposition.

A possible solution to the problem outlined above might be the determination of the number of factors using *fit indices*, whether it be in the context of the EFA model, or preferably, under the EFA within the CFA framework (E/CFA) or exploratory structural equation modeling (ESEM). The E/CFA framework allows the assessment of unrestricted factorial structures with the flexibility of a typical EFA, as the indicators are allowed to load on every factor, while also providing the usual measures of fit and model diagnostics available for the CFA model (Tepper & Hoyle, 1996). In order to specify the E/CFA model each factor has to have 1 marker variable that loads only on that factor (Ferrando & Lorenzo-Seva, 2000). The remaining variables are free to load on every factor. The ESEM framework, which is of very recent development (Asparouhov & Muthén, 2009), has all the benefits of the E/CFA model but does not require the specification of marker variables. In addition, the ESEM model permits the rotation of the factor-loading matrix and can be easily incorporated into larger SEM models.

The new developments in the areas of EFA and SEM modeling in the past 30 years have led methodologists to recommend the use of fit indices for dimensionality assessment, first in the context of the EFA model (e.g., Browne & Cudeck, 1992; Fabrigar et al., 1999; Floyd & Widaman, 1995; Steiger & Lind, 1980), and later, within the E/CFA and ESEM frameworks (e.g., Asparouhov & Muthén, 2009; Hoyle & Duvall, 2004; Ferrando & Lorenzo-Seva, 2000; Tepper & Hoyle, 1996). Furthermore, it has encouraged empirical researchers to use fit indices to establish data dimensionality in the initial – more exploratory – validation phases of their work (e.g., Brown, White, Forsyth, & Barlow, 2004; Campbell-Sills, Liverant, & Brown, 2004; Reichenheim, Moraes, Oliveira, & Lobato, 2011; Sanne, Torsheim, Heiervang, & Stormark, 2009; Tepper & Hoyle, 1996). This synergy between dimensionality and model fit assessment is advantageous because it gives researchers access to important model diagnostic information, such as the presence of correlated errors (Bollen, 2000), when making the determination of the number of factors. It is also valuable because it reduces the need for ad-hoc model manipulation in the more advanced stages of testing, such as the evaluation of a full-blown SEM model, due to a poorly conceived unrestricted structure (Mulaik & Millsap, 2000).

Although fit indices have been recommended and used for dimensionality assessment with increased frequency in recent years, there has been very limited systematic evaluation of their actual efficacy in this area (Fabrigar et al., 1999). With this in mind, the current study is intended to be a preliminary source of information regarding their performance in the dimensionality assessment of ordinal variables, which are commonly found in the social and behavioral sciences. In addition, because the fit of the EFA, E/CFA, and ESEM models will be the same as long as the measurement model has no additional restrictions besides those needed for identification, the results of this study will be equally valid for each of these models.

5.1. Evaluation of Model Fit

Large-sample theory provides a chi-square (χ^2) goodness-of-fit test for comparing a model against a general alternative model based on correlated variables (Bentler & Bonett, 1980). The asymptotic χ^2 statistic tests the null hypothesis that the population covariance matrix of observed variables is equal to the model-implied covariance matrix (Hu & Bentler, 1999). However, because in the social sciences any model is at best an approximation to reality, the null hypothesis of exact fit is known *a priori* to be false, resulting invariably, with a large enough sample, in the rejection of even models that closely approximate the population covariance matrix (Bentler & Bonett, 1980; Browne & Cudeck, 1992; Schermelleh-Engel, Moosbrugger, & Müller, 2003; Yu, 2002).

In order to overcome the limitations of the χ^2 statistical test of exact fit, Bentler and Bonett (1980) introduced the concepts of fit indices and practical significance in relation to the assessment of latent variable models. Bentler and Bonett (1980) proposed the use of *incremental* fit indices as a means to compute the amount of information gained when comparing competing models. They argued that an index of information gained would provide valuable information about the practical usefulness of competing models and should be independent of sample size and statistical significance test information (Bentler & Bonett, 1980). Incremental fit indices assess the degree to which the tested model is superior to an alternative "baseline" model in reproducing the observed covariance matrix. The baseline model is usually a null model in which all the observed variables are uncorrelated (Hu & Bentler, 1999). In addition to the incremental fit indices, measures of absolute fit have also been developed to evaluate the appropriateness of latent variable models. The *absolute* fit indices measure the degree to which the hypothesized model corresponds to the empirical data. In this case, no reference model is used to assess the amount of increment in model fit,

but an implicit comparison is made to a saturated model that exactly reproduces the covariance matrix of observed variables (Hu & Bentler, 1999). Some of the most commonly used fit indices are the Comparative Fit Index (CFI; Bentler, 1990), Tucker-Lewis Index or Non-Normed Fit Index (TLI; Tucker & Lewis, 1973), Root Mean Square Error of Approximation (RMSEA; Steiger & Lind, 1980), and the Standardized Root Mean Square Residual (SRMR; Bentler, 1995). These fit indices, which will be the focus of the current study, have performed relatively well in previous CFA and SEM Monte Carlo studies (e.g., Hu & Bentler, 1999; Sharma, Mukherjee, Kumar, & Dillon, 2005; Yu, 2002), and/or are highly popular in applied research (e.g., Murayama, Elliot, & Yamagata; 2011; Sanne et al., 2009). The formulas for the CFI, TLI, RMSEA, and SRMR indices are given in equations 5.1 to 5.4, respectively.

$$\text{CFI} = 1 - \frac{\max[(\chi_T^2 - df_T), 0]}{\max[(\chi_T^2 - df_T), (\chi_0^2 - df_0), 0]} \tag{5.1}$$

$$\text{TLI} = \frac{\chi_0^2/df_0 - \chi_T^2/df_T}{\chi_0^2/df_0 - 1} \tag{5.2}$$

$$\text{RMSEA} = \sqrt{\max\left\{\frac{\chi_T^2 - df_T}{df_T(N-1)}, 0\right\}} \tag{5.3}$$

$$\text{SRMR} = \sqrt{\left\{2\sum_{i=1}^{p}\sum_{j=1}^{i}\left[\frac{s_{ij} - \hat{\sigma}_{ij}}{s_{ii}s_{jj}}\right]^2\right\}/p(p+1)} \tag{5.4}$$

where χ_T^2 is the chi-square of the hypothesized model, χ_0^2 is the chi-square of the baseline model, df_T are the degrees of freedom of the hypothesized model, df_0 are the degrees of freedom of the baseline model, N is the sample size, s_{ij} is the observed covariance, $\hat{\sigma}_{ij}$ is the

model-implied covariance, s_{ii} and s_{jj} are the observed standard deviations, and p is the number of observed variables.

The CFI is an incremental fit index and has boundaries of 0 and 1, with higher values indicating greater gains in fit in comparison to the baseline model. Likewise, the TLI index is also an incremental fit index that generally ranges from 0 to 1, but, as the index is not normed, it can sometimes obtain values that fall outside of this range. The TLI index differs from the CFI index in that it informs of the relative reduction in misfit *per degree of freedom*, an additional adjustment that takes into account model parsimony (Mahler, 2011). In general, CFI and TLI values of 0.95 or greater are considered to indicate an acceptable fit to the data (Hu & Bentler, 1999; Yu, 2002). The RMSEA index, for its part, is a measure of absolute fit that is concerned with the discrepancy per degree of freedom between the observed covariance matrix and the model-implied covariance matrix. The RMSEA index is bounded below by zero, with lower values indicating a better fit to the data, or what is the same, less error of approximation. Values lower than 0.05 are usually considered as an indication of a close fit to the data according to the RMSEA index (Browne & Cudeck, 1992; Chen, Curran, Bollen, Kirby, & Paxton, 2008; Yu, 2002). Finally, the SRMR index is a measure of absolute fit that computes the average of the standardized residuals between the observed and model-implied covariance matrices. Like the RMSEA index, the SRMR index also has a lower bound of zero, with smaller values indicating a better fit or less residual error. A cutoff value of 0.08 has been recommended for the SRMR index (Hu & Bentler, 1999).

It should be noted that a debate has arisen in recent years on the appropriateness of fixed cutoff values for fit indices in the assessment of latent variable models (e.g., Marsh, Hau, & Wen, 2004; Chen et al., 2008). Some authors have gone as far as to suggest that fit indices be dropped altogether in favor of the traditional statistical tests of exact fit (e.g.,

Barrett, 2007). The criticisms on the use of fixed cutoff values stem from the high Type I and II errors that some indices tend to produce with varying data characteristics (e.g., Mahler, 2011; Nye & Drasgow, 2011). Although these criticisms are valid, no better alternatives seem to be available at the moment, as the determination of optimal cutoff values is dependent partly on population characteristics of the data that are unknown to the researcher and sometimes constitute the very focus of the analysis (e.g., factor loadings, number of factors). A more reasonable alternative may be to try and identify those fit indices that perform relatively well across conditions, while acknowledging the limitations due to the use of fixed cutoff values.

5.2. Dimensionality Assessment of Ordinal Variables

Several important issues need to be taken into consideration when assessing the dimensionality of ordinal-level data. One of these issues is which statistic to use in order to measure the level of association between the ordinal indicators. As it is well known, Pearson's product-moment correlation underestimates the strength of the relationship between ordinal variables (Babakus, Ferguson, & Jöreskog, 1987; Bollen & Barb, 1981), and may produce spurious dimensions known as "difficulty factors" when the variables are skewed in opposite directions (Gorsuch, 1983). Because of these biases, the polychoric correlation coefficient has been recommended as a measure of association for ordinal variables (Flora & Curran, 2004). Assuming that the ordinal variables are a crude measure of underlying bivariate normally distributed variables, polychoric correlations constitute an unbiased estimate of the Pearson correlations between the *underlying* continuous variables (Olsson, 1979a), and have been shown to produce unbiased parameter estimates for both EFA and CFA (Babakus et al., 1987; Flora & Curran, 2004). Despite these advantages, however, there are some lingering concerns about using polychoric correlations for the dimensionality assessment of ordinal variables due to their tendency to produce non-Gramian correlation

matrices (indefinite matrices that have at least one negative eigenvalue) and large sampling errors (e.g., Timmerman & Lorenzo-Seva, 2011; Tran & Formann, 2009; Weng & Cheng, 2005).

Another key issue to consider with ordinal outcome variables is which estimation method to use. One of the most commonly used statistical estimation methods for the common factor model is the maximum likelihood (ML) procedure (Flora & Curran, 2004; Schmitt, 2011). With ordinal variables, however, the assumptions of interval-level measurement and multivariate normality of the ML estimator are violated, leading to biased parameters and standard errors (Beauducel & Herzberg, 2006; Flora & Curran, 2004; Muthén & Kaplan, 1985). A suitable alternative for estimating factor analytic models with ordinal variables and polychoric correlations is robust weighted least squares (Beauducel & Herzberg, 2006; Forero, Maydeu-Olivares, & Gallardo-Pujol, 2009; Yu, 2002), also known as diagonally weighted least squares (DWLS) or weighted least squares with mean- and variance-adjusted standard errors and χ^2 statistic (WLSMV). In general, the WLSMV estimator yields accurate χ^2 statistics, parameter estimates and standard errors across different sample sizes and model complexities (Flora & Curran, 2004; Muthén, du Toit, & Spisic, 1997), and does not require that either the sample polychoric correlation matrix or the asymptotic covariance matrix be Gramian as part of the estimation procedure, an important feature that leads to high convergence rates with small samples and/or many observed variables (Flora & Curran, 2004).

So far, research into the dimensionality assessment of ordinal variables has focused mainly on classic factor retention methods such as Horn's Parallel Analysis or Velicer's Minimum Average Partial Method (e.g., Cho, Li, & Bandalos, 2009; Timmerman & Lorenzo-Seva, 2011). This research has produced mixed findings regarding which measure of association, Pearson or polychoric correlations, leads to better estimates (e.g., Cho et al.,

2009; Tran & Formann, 2009). However, recent studies suggest that polychoric correlations do indeed produce more accurate dimensionality assessments with ordinal indicators, especially as the variables become more skewed (Garrido, Abad, & Ponsoda, 2011, 2012). In general, Parallel Analysis appears to provide accurate assessments of dimensionality across a wide range of conditions (Garrido et al., 2012).

5.3. Determination of the Number of Factors with Fit Indices

Although the performance of fit indices has not been systematically evaluated in the area of dimensionality assessment, the findings from the CFA and SEM literature may give clues as to what can be expected in this area of research. A brief summary of the most relevant findings from this literature is given below.

First, the χ^2 statistic has shown a tendency to overestimate its theoretical value with small samples and large numbers of observed variables, leading to poor fit values and high Type I errors (Hu & Bentler, 1999; Kenny & McCoach, 2003; Sharma et al., 2005; Yu, 2002). These results suggest that the χ^2 based fit indices are likely to overestimate the numbers of factors in these conditions. *Second*, moderate to high levels of skewness have also led to inflated Type I error rates for the χ^2 statistic and the CFI, TLI, RMSEA, and SRMR fit indices, when the sample size was small (Lei, 2009; Savalei & Rhemtulla, in press; Yu, 2002). Therefore, the fit indices may tend to overestimate the number of factors when the ordinal variables are skewed and the sample size is small. *Third*, decisions based on the CFI and TLI indices have been highly similar, leading to agreement rates of approximately 0.99 for both correctly specified and misspecified models (Yu, 2002). Additionally, these indices have performed particularly well with CFA structures in detecting an incorrect specification of the number of latent factors (Mahler, 2011). Based on these findings, the CFI and TLI indices can be expected to produce similar, as well as relatively accurate, dimensionality

estimates. *Fourth*, the RMSEA index has shown an extreme sensitivity to the size of the factor loadings, producing very high Type II errors with low factor loadings and very high Type I errors with exceptionally high factor loadings (Browne, MacCallum, Kim, Andersen, & Glaser, 2002; Heene, Hilbert, Draxler, Ziegler, & Bühner, 2011; Mahler, 2011). According to these results, the RMSEA index can be expected to strongly underestimate the number of factors with low factor loadings (exceptionally high factor loadings [e.g., 0.90] were not simulated in the current study). *Last*, the SRMR index has performed erratically with ordinal variables, generating high Type I errors with small samples and high Type II errors with large samples (Yu, 2002). In light of these findings, the SRMR index is likely to underfactor and overfactor with small and large samples, respectively.

5.4. Factor Retention Methods Evaluated in the Current Study

The current study evaluated the performance of 4 commonly used fit indices in determining the number of factors with ordinal variables. The selected fit indices were comprised of 2 incremental fit indices, the Comparative Fit Index (**CFI**; Bentler, 1990) and the Tucker-Lewis Index (**TLI**; Tucker & Lewis, 1973), and 2 absolute fit indices, the Root Mean Square Error of Approximation (**RMSEA**; Steiger & Lind, 1980) and the Standardized Root Mean Square Residual (**SRMR**; Bentler, 1995). In addition to these fit indices, 3 classic factor retention methods developed in the framework of PCA extraction were also evaluated in order to compare and give context to the performance of the fit indices: the eigenvalue-greater-than-1 criterion or Kaiser's rule (**K1**; Kaiser, 1960), Parallel Analysis (**PA**; Horn, 1965), and the Minimum Average Partial Method (**MAP**; Velicer, 1976). In terms of the expected performance of the classic factor retention methods, extensive research with continuous and ordinal variables has shown that PA is highly accurate, while K1 grossly overestimates the numbers of factors and MAP is strongly biased toward underfactoring with low factor loadings and/or a small number of variables per factor (e.g., Garrido et al., 2011,

2012; Timmerman & Lorenzo-Seva, 2011; Velicer et al., 2000; Zwick & Velicer, 1986). These 3 methods are expected to provide high, low, and medium levels of accuracy, respectively, by which to judge and compare the performance of the fit indices. A brief description of each classic factor retention method is given below.

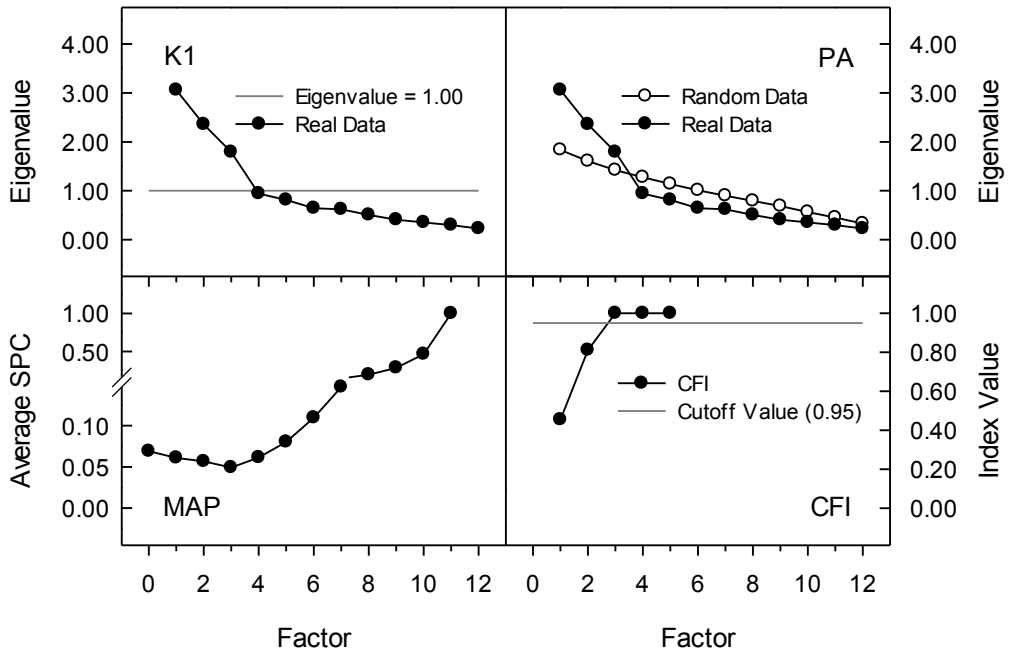
The K1 rule posits that only factors with eigenvalues > 1 should be retained. This criterion is based on Kaiser's (1960) proofs and arguments that Guttman's (1954) latent-root-one lower bound estimate for the minimum rank of a correlation matrix could be used as a psychometric upper bound for the number of factors problem. The rationale behind this rule is that a factor should be able to explain at least as much variance as a variable is accorded in the standard score space (Dickman, 1960), and that a threshold of 1 ensures that the component will have a positive internal consistency (Kaiser, 1960).

The MAP method was also developed in the context of PCA and is based on the matrix of partial correlations. Each component is partialled out of the correlation matrix and the average of the squared partial correlations is computed. The number of factors to retain is determined by the point where the minimum average of the squared partial correlations is obtained. The rationale of this procedure can be described as follows: as common variance is partialled out of the correlation matrix for each successive component, the MAP criterion will keep on decreasing. At the point where the common variance has been removed, extracting additional components will result in unique variance being partialled out, and the MAP criterion will begin to rise. The MAP procedure, therefore, provides an unequivocal stopping point for the number of factors by separating the common and unique variance and retaining only those factors that consist primarily of common variance (Velicer et al., 2000). For a thorough presentation of the procedure used to compute the MAP method with ordinal variables see Garrido et al. (2011).

Parallel Analysis was the final factor retention method included in this study. Horn (1965) argued that because the proofs for the K1 rule were performed on population statistics, some components from uncorrelated variables in the population could have eigenvalues > 1 at the sample level due to sampling error, and least squares "capitalization" on this error in the computation of the latent roots. Therefore, he proposed PA as a means to estimate and take into account the proportion of variance that was due to sampling error and chance capitalization. In this sense, PA may be viewed as a sample alternative to the K1 rule. Instead of retaining factors that have eigenvalues > 1, with PA only those factors that have eigenvalues greater than those generated from independent variates are retained. The goal is to account for chance capitalization in the sample eigenvalues under the null hypothesis of independent variables (Buja & Eyuboglu, 1992). For a detailed description on how to perform PA with ordinal variables see Garrido et al. (2012).

A visual representation of the type of results that may be obtained by the factor retention methods is shown in Figure 5.1. All the methods were applied to the same data matrix, which was sampled from a robust 3-factor population structure in order to obtain the correct dimensionality estimate with all the factor retention methods. Note than in practice the methods will not necessarily agree on the same number of factors, and may in fact, often disagree.

The procedure illustrated for the CFI index in Figure 5.1 is similar to the one that would be followed for the remaining fit indices. That is, a 1-factor model is initially fitted to the correlation matrix and if the index shows a lack of it, more factors are sequentially extracted until an acceptable level of fit is reached. The number of factors is determined at the point where an acceptable fit is first obtained. In this case, the CFI index had values of 0.45, 0.81, and 1.00, for unrestricted structures of 1 factor, 2 factors, and 3 factors, respectively. So, with a cutoff value of 0.95, the CFI would suggest 3 factors, the correct number in this case.



Note. SPC = squared partial correlation; all the methods were applied to the same dataset that was sampled from a 3-factor population structure.

Figure 5.1: *Graphic Examples of the Different Factor Retention Methods*

The determination of the number of factors with the classic factor retention methods of K1, PA, and MAP, can also be seen in Figure 5.1. In the case of the K1 rule, the eigenvalues for the first 4 factors were 3.06, 2.36, 1.79, and 0.94, respectively. Because only the first three factors had eigenvalues > 1, a 3-factor solution is suggested. The interpretation of PA is similar to that of K1, with the exception that the criterion eigenvalues instead of being a constant of 1 are a variable number depending on the number of the factor. The criterion eigenvalues for PA (the mean eigenvalues across 100 random samples) were 1.83, 1.61, 1.42, and 1.27, for the first 4 factors, respectively. In this case, the PA method would also suggest a 3-factor structure as each of the first 3 factors had an eigenvalue that was higher than the mean of its random counterpart, but the 4th factor, on the contrary, had an eigenvalue of 0.94 that was lower than the eigenvalue of 1.27 obtained for the 4th factor of the random datasets. Lastly, the MAP criterion was 0.0691, 0.0609, 0.0567, 0.0494, and 0.0614 for structures of 0

to 4 factors, respectively (the values continued to increase after the 4th factor). Because the minimum MAP value of 0.0494 was obtained for 3 factors, that is, after partialling out the first 3 components of the correlation matrix, the MAP criterion estimated 3 factors as well.

5.5. Goals of the Current Study

The main goal of the current study is to evaluate the efficacy of 4 fit indices, CFI, TLI, RMSEA, and SRMR, in determining the number of factors with observed ordinal variables. In order to give context to the performance of these fit indices and to gauge their comparative accuracy, 3 classic factor retention methods developed in the framework of PCA extraction, K1, MAP, and PA, were also included in the simulation design. A secondary goal is to establish an optimal cutoff value for each fit index according to their performance in the dimensionality assessment of ordinal variables. An index value is considered to be optimal, or near optimal, if it produces a relatively high proportion of correct dimensionality estimates along with a relatively small error of estimation.

5.6. Method

5.6.1. Design

The performance of the factor retention methods was assessed through a simulation study using Monte Carlo methods. The factorial design included the manipulation of 4 "structure" factors – factor loading, number of variables per factor, number of factors, and factor correlation – and 3 "sample" factors – sample size, skewness, and number of response categories – for a total of seven independent variables. Altogether, these 7 variables have been shown to affect the performance of factor retention methods with ordinal and/or continuous variables (Cho et al., 2009; Garrido et al., 2011, 2012; Timmerman & Lorenzo-

Seva, 2011; Velicer et al, 2000; Zwick & Velicer, 1986). A summary of the research design is presented in Table 5.1.

Table 5.1

Research Design

	Level			
Independent Variables	L1	L2	L3	L4
Structure Factors				
Factor Loading	0.40	0.55	0.70	
Variables per Factor	4	8	12	
Number of Factors	1	2	4	
Factor Correlation	0.00	0.30	0.50	
Sample Factors				
Sample Size	100	300	1000	
Skewness	0	±1	±2	
Response Categories	2	3	4	5

Note. L = level.

The factorial design shown in Table 5.1 can be broken up into two parts: (1) the unidimensional condition with a 3 x 3 x 3 x 3 x 4 (factor loading x variables per factor x sample size x skewness x response categories) design, for a total of 324 factor combinations; and (2) the multidimensional condition with a 3 x 3 x 2 x 3 x 3 x 3 x 4 (factor loading x variables per factor x number of factors x factor correlation x sample size x skewness x response categories) design, for a total of 1,944 factor combinations. In all, 2,268 factor combinations were studied.

The levels for the independent variables were chosen so that they were representative of the range of values that are encountered in applied settings. In each case, an attempt was made to include a small/weak, medium/moderate, and large/strong level. For instance, according to Comrey and Lee (1992) sample sizes of 100, 300, and 1000 can be considered as poor, good, and excellent. Similarly, these authors consider factor loadings of 0.40, 0.55, and

0.70, to be poor, good and excellent as well. For the factor correlations, the orthogonal condition (r = 0.00) was included, plus moderate (r = 0.30), and strong (r = 0.50) correlation levels, according to Cohen's (1988) criterion. Additionally, 4 variables per factor has been recommended as the minimum level of identification that a factor should have (Mulaik & Millsap, 2000), 8 can be considered as a moderately strong factor (Velicer et al., 2000), and 12 as a highly overidentified factor (Widaman, 1993). Furthermore, 5 was chosen as the maximum number of response categories to be simulated because gains in reliability and validity appear to be only marginal with more scale points (Preston & Colman, 2000). In terms of the skewness of the ordinal variables, 0 indicates a symmetrical distribution, whereas ±1 may be considered as a meaningful departure from normality (Meyers, Gamst, & Guarino, 2006, p. 50), and ±2 as a high level of skewness (Muthén & Kaplan, 1985). Finally, structures of 1, 2, and 4 factors were simulated, which includes the unidimensional condition, as well as relatively low to medium values for modern multidimensional inventories. It should be noted that it was not possible to simulate structures of more than 4 factors because of the large research design and the rapid increase in computational time with higher numbers of observed variables.

Regarding the factor retention methods, 4 fit indices were studied: CFI, TLI, RMSEA, and SRMR. Because the performance of these fit indices had not been systematically studied in the area of dimensionality assessment, a wide range of cutoff values were evaluated. For the CFI and TLI indices, 20 different cutoff values were evaluated from 0.80 to 0.99 in increments of 0.01. Similarly, for the RMSEA and SRMR indices, 20 cutoff values were evaluated from 0.20 to 0.01 in decrements of 0.01. The procedure used to determine the number of factors with the fit indices started by fitting a 1-factor model and comparing its fit to the prespecified cutoff value of the fit index; if the model fit acceptably, the index suggested a 1-factor solution, if not, the number of factors was sequentially increased by 1

until a model with an acceptable fit was obtained. All the factor analyses were performed using methods developed to accommodate ordinal variables. Specifically, the common factor model was fitted to polychoric correlations (Olsson, 1979a) with the WLSMV estimator (Muthén et al., 1997).

The performance of the classic factor retention methods K1, MAP, and PA, was also evaluated in the current study. The implementation of the MAP method followed the descriptions and recommendations in Garrido et al. (2011), which include using polychoric correlations, squaring the partial correlations and smoothing the non-Gramian polychoric matrices with the "ridge" procedure. Similarly, the PA method was computed in accordance to the recommendations and simulation procedures in Garrido et al. (2012), which include using polychoric correlations, PCA extraction, and the mean eigenvalue criterion. Also, the eigenvalues from the non-Gramian polychoric matrices were not given any treatment, as the performance of the PA method appears to be unaffected by the emergence of these type of matrices (Garrido et al., 2012). This latter consideration was applied to the K1 rule as well.

5.6.2. Data Generation

For each of the 2,268 factor combinations, 100 sample data matrices of ordinal variables were generated according to the following common factor model procedure: first, the reproduced population correlation matrix (with communalities in the diagonal) is computed

$$\mathbf{R}_{\mathrm{R}} = \boldsymbol{\Lambda}\boldsymbol{\phi}\boldsymbol{\Lambda}^{\mathrm{T}} \tag{5.5}$$

where $\mathbf{R}_{\mathrm{R}}$ is the reproduced population correlation matrix, $\boldsymbol{\Lambda}$ is the population factor loading matrix, and $\boldsymbol{\phi}$ is the population factor correlation matrix.

The population correlation matrix $\mathbf{R}_P$ is then obtained by inserting unities in the diagonal of $\mathbf{R}_R$, thereby raising the matrix to full rank. The next step is performing a Cholesky decomposition of $\mathbf{R}_P$, such that

$$\mathbf{R}_P = \mathbf{U}^T\mathbf{U} \tag{5.6}$$

where **U** is an upper triangular matrix.

The sample matrix of continuous variables **X** is subsequently computed

$$\mathbf{X} = \mathbf{ZU} \tag{5.7}$$

where **Z** is a matrix of random standard normal deviates with rows equal to the sample size and columns equal to the number of variables.

The sample matrix of ordinal variables is obtained by applying a set of thresholds to **X** according to the specified levels of skewness and number of response categories (for a list of the thresholds used in this study see Garrido et al., 2011). The thresholds for the symmetric condition (skewness = 0) were computed by partitioning the continuum from z = -3 to z = 3 at equal intervals (see Bollen & Barb, 1981). Thresholds for the asymmetric conditions were created so that as the skewness level increased, the observations were "piled up" in one of the extreme categories (see Muthén & Kaplan, 1985). In addition, half of the variables of each factor were categorized with the same positive skewness and the other half with the same negative skewness.

5.6.3. Non-Convergence

When conducting the factor analyses it was necessary to determine how to best deal with solutions that did not converge. An initial decision was made that any method proposed to deal with non-convergence should meet the requirement of being able to be implemented

in practice, as well as in simulation studies. For this reason, the option of screening out and/or re-simulating the cases with non-converged solutions was discarded, as this possibility is not available in a practical scenario. On the other hand, because non-convergence often results from overfactoring (Hoyle & Duvall, 2004; Raykov & Marcoulides, 2011, p.68), the decision was made to treat it as a diagnostic of *overfitting* (Fabrigar et al., 1999) and a signal to stop the factor extraction process. Operationally, when non-convergence occurred the number of factors for those fit indices and cutoff values that had yet to be satisfied was fixed at the number extracted in the previous converged solution. In order to assess the appropriateness of this procedure, the solutions that did not converge were flagged and their numbers of factors were coded. The results revealed that non-convergence occurred in 33,048 cases (14.6% of the total number of cases)[3], of which 32,132 (97.2%) were overextractions, 730 (2.2%) were for the same number of factors as those present in the population, and 186 (0.6%) were underextractions. The fact that the overwhelming majority of solutions that did not converge (97.2%) occurred for overfitted models supports the interpretation of non-convergence as attempts to estimate unrestricted models with too many factors.

5.6.4. Assessment Criteria

The accuracy of the factor retention methods was evaluated according to three complementary criteria: the proportion of correct estimates (PC), the mean bias error (MBE), and the mean absolute error (MAE). The corresponding formula for each criterion is presented in Equations 5.8 to 5.10:

$$PC = \frac{\sum C}{N_S}, \quad \text{for} \quad C = \begin{cases} 1 & \text{if } \hat{\theta} = \theta \\ 0 & \text{if } \hat{\theta} \neq \theta \end{cases} \tag{5.8}$$

[3] The large number of solutions that did not converge was due to the inclusion of extremely stringent fit criteria as part of the attempt to evaluate the performance of the fit indices across a broad array of cutoff values.

$$\mathrm{MBE} = \frac{\sum(\hat{\theta} - \theta)}{N_S} \tag{5.9}$$

$$\mathrm{MAE} = \frac{\sum |\hat{\theta} - \theta|}{N_S} \tag{5.10}$$

where N_S is the number of sample data matrices generated for each factor combination (100), $\hat{\theta}$ is the estimated number of factors, and θ is the population number of factors.

The PC criterion has boundaries of 0 and 1, with 0 indicating a total lack of accuracy and 1 reflecting perfect accuracy. In contrast, a 0 on the MBE criterion shows a complete lack of bias, with negative and positive values indicating underfactoring and overfactoring, respectively. It is important to note that the MBE cannot be used alone as a measure of method performance because errors of under- and overfactoring can compensate each other and give a false illusion of accuracy (this does not happen with the PC or MAE criteria). In terms of the MAE criterion, higher values signal larger deviations from the population number of factors, while a value of 0 indicates perfect accuracy. The MAE criterion was chosen over the root mean square error (RMSE), which is a more frequently used measure of average model-performance error, because it is only a function of the average-error magnitude whereas the RMSE also varies as a function of the square root of the number of errors and with the variability within the distribution of the error magnitudes, making it more difficult to interpret and to compare across studies (Willmott & Matsuura, 2005). The MAE, on the other hand, is a straightforward statistic that can be interpreted for this particular study as the average number of factors by which the estimates of the factor retention methods deviate from the population number of factors. This information is particularly useful to researchers, who, when making the final decision on the number of factors to retain, usually take into account the interpretability of the rotated factor solutions adjacent or in the

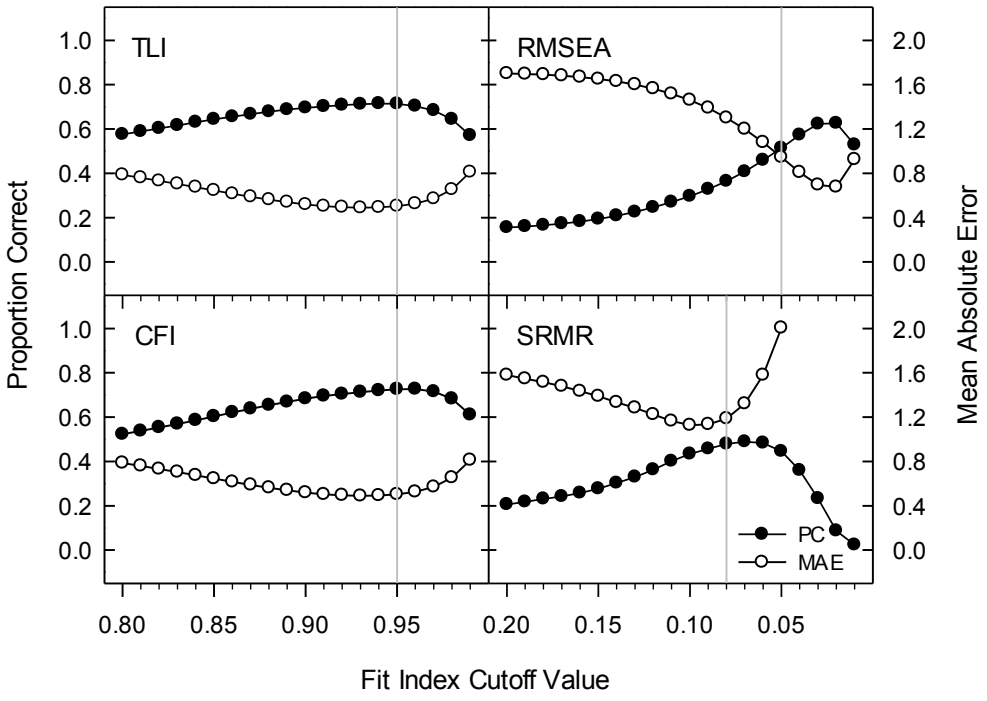
proximity to the one(s) indicated by the factor retention method(s) (e.g., Brown et al., 2004; Campbell-Sills et al., 2004).

All simulations were run under the MATLAB program (version R2010a; The MathWorks, Inc., 1984-2010). The values for the fit indices were obtained from the M*plus* program (version 6.11; Muthén & Muthén, 1998-2010) and were later inputted into the MATLAB environment. In addition, the classic factor retention methods were programmed directly into MATLAB code by the authors. In all cases, the polychoric correlations were computed using the maximum likelihood algorithms provided by Olsson (1979a).

5.7. Results

A look at the performance of the fit indices across the different cutoff values (cv) is presented in Figure 5.2. The results indicate that the accuracy of the CFI, TLI, and SRMR indices was at its best, or near best, at the recommended cutoff values of 0.95, 0.95, and 0.08, respectively. In the case of the CFI index, at the 0.95cv the PC was 0.73, its highest overall PC, while the MAE was 0.50, a value approximately equal to the lowest MAE of 0.49, obtained at the 0.93cv. Similarly, at the 0.95cv the TLI index had a PC of 0.71, its highest overall PC, and a MAE of 0.57, a marginally higher level of error than the overall lowest MAE of 0.53, obtained at the 0.92cv. Regarding the SRMR index, at the 0.08cv the PC and MAE were 0.48 and 1.19, respectively, very similar to the overall highest PC of 0.49, obtained at the 0.07cv, and the overall lowest MAE of 1.13, obtained at the 0.10cv. A particular characteristic of the SRMR index was that beyond the 0.08cv it lost accuracy very rapidly (according to the MAE criterion) as the cutoff values became more stringent. For this reason, cutoff values lower than 0.08 are clearly not recommended for the SRMR index. In contrast to the other fit indices, the RMSEA index behaved poorly at the recommended cutoff value of 0.05. Specifically, at the 0.05cv the PC and MAE were 0.51 and 0.94, respectively, a

notably poorer performance than the highest PC of 0.63 and the lowest MAE of 0.68, obtained for the 0.02cv. Also of note for the RMSEA index was that the performance changed substantially in the critical range of 0.10 to 0.01 with small changes in the cutoff values (e.g., changes of 0.01), something that did not occur for the other χ^2 based indices (CFI and TLI) at an analogous range of 0.90 to 0.99.



Note. PC = proportion correct; MAE = mean absolute error; the MAE criterion is omitted for SRMR ≤ 0.04 in order to facilitate the comparison of the fit indices' performance (the omitted MAE values are: MAE[SRMR$_{0.04}$] = 2.67; MAE[SRMR$_{0.03}$] = 3.63; MAE[SRMR$_{0.02}$] = 4.83; MAE[SRMR$_{0.01}$] = 5.73); the grey vertical lines indicate the most recommended cutoff values from the confirmatory factor analysis and structural equation modeling literature.

Figure 5.2: *Performance of the Fit Indices across successive Cutoff Values*

Given the optimal performance of the CFI, TLI, and SRMR indices at the recommended cutoff values of 0.95, 0.95, and 0.08, respectively, the rest of the results for these indices will be presented only for these cutoff values. On the other hand, because the

RMSEA index performed poorly at the recommended cutoff value of 0.05, the rest of the results for this index will be presented for both the recommended 0.05cv and the optimal 0.02cv.

In order to evaluate the correspondence amongst the different dimensionality estimates, the Pearson correlation was computed between the numbers of factors suggested by the factor retention methods across the 226,800 data matrices that were simulated (see Table 5.2). The results in Table 5.2 show that the estimates between the $CFI_{0.95}$ and $TLI_{0.95}$ indices had nearly identical correspondence, with a correlation of 0.98. Additionally, the $CFI_{0.95}$ and $TLI_{0.95}$ indices had a high level of correspondence with $RMSEA_{0.02}$, attaining correlations of 0.87 and 0.88, respectively. In contrast, the $RMSEA_{0.05}$ index only had a marginal level of correspondence with the other fit indices ($r \leq 0.38$), and only a moderate level with $RMSEA_{0.02}$ ($r = 0.52$). The final fit index, $SRMR_{0.08}$, showed moderate correspondence with $CFI_{0.95}$ and $TLI_{0.95}$ ($r = 0.73$), as well as $RMSEA_{0.02}$ ($r = 0.68$).

Table 5.2

Correlations between the Numbers of Factors estimated by the Factor Retention Methods

Method	1	2	3	4	5	6	7	8
1. $CFI_{0.95}$	1.00							
2. $TLI_{0.95}$	0.98	1.00						
3. $RMSEA_{0.02}$	0.87	0.88	1.00					
4. $RMSEA_{0.05}$	0.38	0.36	0.52	1.00				
5. $SRMR_{0.08}$	0.73	0.73	0.68	0.28	1.00			
6. K1	0.67	0.67	0.48	-0.05	0.61	1.00		
7. MAP	0.32	0.29	0.35	0.40	0.19	0.35	1.00	
8. PA	0.70	0.68	0.59	0.36	0.43	0.60	0.44	1.00

Note. The numbers next to the fit indices indicate their cutoff value; $p < 0.01$ for all cases.

Regarding the classic factor retention methods, K1 showed moderate correspondence with $CFI_{0.95}$ and $TLI_{0.95}$ ($r = 0.67$), $SRMR_{0.08}$ ($r = 0.61$), and PA ($r = 0.60$). The moderate correspondence between K1 and PA was expected as both methods are based on the size of

the eigenvalues and are equivalent at the population level. However, the moderate correspondence with $CFI_{0.95}$, $TLI_{0.95}$, and $SRMR_{0.08}$ was surprising, and, as will be seen later, was due to the overestimation of the number of factors by all of these methods with small samples. The MAP method, on the other hand, had low levels of correspondence with each of the other 7 factor retention methods ($r \leq 0.44$). This result was expected because the MAP method is particular for greatly underestimating the number of factors with low factor loadings and/or a small number of variables per factor. Finally, the PA method showed moderate correspondence with $CFI_{0.95}$ ($r = 0.70$), $TLI_{0.95}$ ($r = 0.68$), K1 ($r = 0.60$), and $RMSEA_{0.02}$ ($r = 0.59$), and low to marginal correspondence with MAP ($r = 0.44$), $SRMR_{0.08}$ ($r = 0.43$), and $RMSEA_{0.05}$ ($r = 0.36$).

An evaluation of the overall performance of the factor retention methods is presented next in Table 5.3. Among the fit indices, the $CFI_{0.95}$ index was the most accurate (PC = 0.73), had the smallest bias (MBE = 0.21), and produced the smallest error of estimation (MAE = 0.50). As expected due to their high level of correspondence, the $TLI_{0.95}$ index had a very similar, although slightly poorer, performance in comparison to the $CFI_{0.95}$ index (PC = 0.71; MBE = 0.34; MAE = 0.57). Next in line was the $RMSEA_{0.02}$ index, with a performance that was a step below that of $CFI_{0.95}$ and $TLI_{0.95}$ ($PC = 0.63 < 0.71 < 0.73$; $MAE = 0.68 > 0.57 > 0.50$). In general, these three indices estimated the number of factors correctly for more than half of the data matrices ($PC \geq 0.63$), showed a small tendency to overfactor ($0.21 \leq MBE \leq 0.34$), and their estimates were, on average, less than 1 factor apart from the population number of factors ($MAE \leq 0.68$). The last two fit indices, $RMSEA_{0.05}$ and $SRMR_{0.08}$, performed noticeably poorer than the $CFI_{0.95}$, $TLI_{0.95}$, and $RMSEA_{0.02}$ indices. Apart from the $RMSEA_{0.05}$ index being less accurate ($PC = 0.51 < 0.63$) and producing a larger error of estimation ($MAE = 0.94 > 0.68$) than its counterpart $RMSEA_{0.02}$, it showed a strong tendency to underfactor (MBE = -0.76), becoming the only fit index to exhibit this behavior. The

$SRMR_{0.08}$ index, for its part, had the worst performance of all the fit indices, producing the least accurate estimations (PC = 0.48), and the largest error of estimation (MAE = 1.19).

The performances of the fit indices can be compared and put into context by examining the performances of the classic factor retention methods K1, MAP, and PA. As can be seen in Table 5.3, K1 was clearly the worst performing of the 8 factor retention methods (PC = 0.32, lowest accuracy; MAE = 3.53, largest error), highlighted by its gross overestimation of the number of factors (MBE = 3.53). The MAP method, although moderately more accurate than $RMSEA_{0.05}$ (PC = 0.58 > 0.51), had a comparable error of estimation (MAE = 0.91 ≈ 0.94) and an analogous tendency to underfactor (MBE = -0.91). Both the MAP method and the K1 rule emerged as extremely biased methods, as evidenced by the fact that they *always* erred on the same direction (|MBE| = MAE). Finally, PA had the best performance of all the factor retention methods, with the highest accuracy (PC = 0.81), smallest error of estimation (MAE = 0.31), and a negligible level of bias (MBE = -0.02). Overall, the performance level of the three-best fit indices ($CFI_{0.95}$, $TLI_{0.95}$, and $RMSEA_{0.02}$) was somewhere in between the performances of PA and MAP.

Table 5.3

Overall Method Performance

	Performance Criteria		
Method	PC	MBE	MAE
$CFI_{0.95}$	0.73	0.21	0.50
$TLI_{0.95}$	0.71	0.34	0.57
$RMSEA_{0.02}$	0.63	0.26	0.68
$RMSEA_{0.05}$	0.51	-0.76	0.94
$SRMR_{0.08}$	0.48	0.24	1.19
K1	0.32	3.53	3.53
MAP	0.58	-0.91	0.91
PA	0.81	-0.02	0.31

Note. PC = proportion correct; MBE = mean bias error; MAE = mean absolute error; the numbers next to the fit indices indicate their cutoff values.

With the purpose of gauging the sensitivity of the factor retention methods to the independent variables, an analysis of variance (ANOVA) was performed separately for each method using the MAE statistic as the dependent variable. Although preliminary analyses showed that the MAE and PC statistics produced similar ANOVA results, the MAE criterion led to the identification of a greater number of relevant interactions. Unlike the PC criterion, which has an upper bound of 1, the MAE statistic can theoretically obtain any value above 0, thus allowing for a wider distribution of values and more Power. Also, because of the large sample size, most of the effects were expected to be significant. For this reason, the partial omega square (ω_p^2) measure of effect size was chosen to establish the impact of the independent variables and their interactions. According to Cohen (1988), values of 0.01 represent small effects, 0.06 medium effects, and 0.14 or more, large effects.

The results of the ANOVAs can be seen in Table 5.4. In order to facilitate the interpretation of these results, their commentary will be divided into two parts. In the first part, the main effects of the independent variables will be analyzed in combination with the performance of the factor retention methods across the different factor levels (Tables 5.5 and 5.6). In the second part, the most salient interactions will be discussed with the help of Figures 5.3 to 5.5. Each interaction will be represented by an individual Figure and will include the factor retention methods that were most affected by it.

A look at Table 5.4 reveals that the $CFI_{0.95}$, $TLI_{0.95}$, and $RMSEA_{0.02}$ indices were affected similarly by the structure and sample factors. The most salient variable for these fit indices was the sample size, with large effects sizes for all three indices ($\omega_p^2 \geq 0.30$). With small samples of 100 observations, these indices exhibited their worst overall performance, including their smallest PC (e.g., [Table 5.6]: $PC[CFI_{0.95}] = 0.50$, lowest overall PC for the $CFI_{0.95}$ index) and largest MAE (e.g., [Table 5.6]: $MAE[TLI_{0.95}] = 1.30$, highest overall MAE for the $TLI_{0.95}$ index), while strongly overestimating the number of factors (e.g., [Table

5.6]: MBE[RMSEA$_{0.02}$] = 0.95). On the other hand, with medium sample sizes of 300, their performance improved dramatically (e.g., [Table 5.6]: with sample sizes of 300, MAE[TLI$_{0.95}$] = 0.32 << 1.30 = MAE[TLI$_{0.95}$] with sample sizes of 100), and with large sample sizes of 1000, it became highly accurate (e.g., [Table 5.6]: PC[CFI$_{0.95}$] = 0.90).

Table 5.4

ANOVA Univariate Effect Sizes for the Independent Variables and their Interactions

	Method							
Independent Variables	CFI (0.95)	TLI (0.95)	RMSEA (0.02)	RMSEA (0.05)	SRMR (0.08)	K1	MAP	PA
Population Factors								
FLOAD	0.13	0.12	0.10	0.65	0.45	0.90	0.71	0.18
VARFAC	0.06	0.06	0.04	0.06	0.35	0.95	0.70	0.02
FAC	0.13	0.10	0.08	0.72	0.52	0.89	0.53	0.07
FACCORR	0.01	0.00	0.01	0.19	0.07	0.00	0.04	0.02
Sample Factors								
N	0.34	0.35	0.30	0.05	0.49	0.83	0.23	0.18
SKEW	0.17	0.15	0.18	0.14	0.35	0.57	0.14	0.05
RESCAT	0.01	0.00	0.00	0.09	0.17	0.37	0.04	0.02
Interactions								
VARFAC * FAC	0.01	0.02	0.00	0.03	0.02	0.76	0.18	0.01
FLOAD * VARFAC	0.02	0.02	0.00	0.01	0.05	0.71	0.38	0.00
FLOAD * FAC	0.03	0.02	0.08	0.42	0.13	0.47	0.23	0.03
FLOAD * N	0.05	0.05	0.01	0.09	0.04	0.31	0.01	0.07
FAC * N	0.07	0.08	0.01	0.01	0.03	0.27	0.07	0.03
VARFAC * SKEW	0.08	0.07	0.09	0.01	0.26	0.27	0.00	0.00
FLOAD * VARFAC * N	0.05	0.05	0.01	0.00	0.10	0.26	0.14	0.01
FLOAD * FAC * N	0.01	0.01	0.01	0.01	0.01	0.17	0.00	0.00
FLOAD * VARFAC * FAC	0.00	0.01	0.00	0.01	0.01	0.15	0.08	0.00
FLOAD * SKEW * N	0.01	0.01	0.02	0.08	0.01	0.15	0.01	0.00
VARFAC * N	0.17	0.16	0.10	0.01	0.53	0.56	0.01	0.00
SKEW * N	0.18	0.16	0.15	0.00	0.51	0.01	0.09	0.02
VARFAC * SKEW * N	0.15	0.12	0.13	0.01	0.25	0.01	0.00	0.00
RESCAT * N	0.03	0.02	0.04	0.01	0.24	0.02	0.01	0.01
RESCAT * SKEW * N	0.07	0.06	0.06	0.02	0.20	0.02	0.01	0.00
VARFAC * RESCAT	0.02	0.01	0.01	0.00	0.15	0.13	0.00	0.00

Note. FLOAD = factor loading; VARFAC = variables per factor; FAC = number of factors; FACCORR = factor correlation; N = sample size; RESCAT = number of response categories; SKEW = skewness; the numbers in parenthesis indicate the cutoff values of the fit indexes; the effect sizes are measured by the partial omega squared statistic (ω_p^2); large effect sizes ($\omega_p^2 \geq 0.14$) are highlighted in grey shade; the dependent variable is the mean absolute error statistic.

The other independent variable that reached a large effect size for the $CFI_{0.95}$, $TLI_{0.95}$, and $RMSEA_{0.02}$ fit indices was the skewness of the ordinal variables ($\omega_p^2 \geq 0.15$). As can be seen in Table 5.6, the accuracy levels with skewness of 0 and ±1 were similar (e.g., with skewness of 0, $PC[TLI_{0.95}] = 0.79 \approx 0.77 = PC[TLI_{0.95}]$ with skewness of ±1), but dropped off substantially with skewness of ±2 (e.g., $PC[TLI_{0.95}] = 0.58$), where the fit indices showed a marked tendency overfactor (e.g., $MBE[TLI_{0.95}] = 0.74$).

Next in line for the $CFI_{0.95}$, $TLI_{0.95}$, and $RMSEA_{0.02}$ fit indices were the factor loading and number of factors variables, with effects of medium size ($0.08 < \omega_p^2 < 0.13$). The impact of both of these variables were in accordance with the theoretical expectations, with the performance improving as the factor loadings increased (e.g., [Table 5.5]: $PC[CFI_{0.95}] = 0.56$, 0.76, and 0.86, with factor loadings of 0.40, 0.55, and 0.70, respectively) and deteriorating as the number of factors increased (e.g., [Table 5.5]: $MAE[TLI_{0.95}] = 0.21$, 0.37, and 0.89, with 1, 2, and 4 factors, respectively). It should be noted that the performance of the $RMSEA_{0.02}$ index was especially poor for the 1-factor structure in comparison to the other factor retention methods (e.g., [Table 5.5]: $PC[RMSEA_{0.02}] = 0.68$, lowest PC for all methods except K1), where the index tended to moderately overfactor (e.g., [Table 5.5]: $MBE[RMSEA_{0.02}] = 0.41$). In contrast to the performances across the levels of factor loading and number of factors, the errors of estimations of these fit indices actually became greater with more variables per factors (e.g., [Table 5.5]: $MAE[CFI_{0.95}] = 0.37$, 0.43, and 0.71, with 4, 8, and 12 variables per factor, respectively), conditions that are expected to be more robust and have clearer dimensionality. While undesirable, this result was not surprising, as larger numbers of variables per factor imply larger numbers of total variables and free model parameters, and consequently, a need for larger sample sizes in order to obtain accurate estimates and fit statistics. Finally, the factor correlation and number of response categories had near-zero effect sizes ($\omega_p^2 \leq 0.01$), indicating that these variables had a minimal impact in the errors of

estimation of the above-mentioned fit indices (e.g., [Table 5.5]: MAE[$TLI_{0.95}$] = 0.63, 0.59, and 0.68, with factor correlations of 0.00, 0.30, and 0.50, respectively).

The remaining two fit indices, $RMSEA_{0.05}$ and $SRMR_{0.08}$, exhibited very erratic and biased performances. According to the results in Table 5.4, The $RMSEA_{0.05}$ index was extremely sensitive to the number of factors ($\omega_p^2 = 0.72$) and the factor loadings ($\omega_p^2 = 0.65$). In general, the $RMSEA_{0.05}$ index tended to greatly underfactor with low factor loadings and/or a high number of factors (e.g., [Table 5.5]: with 4 factors, MBE[$RMSEA_{0.05}$] = -1.66), while producing some of the least accurate estimations of all fit indices (e.g., [Table 5.5]: with factor loadings of 0.40, PC[$RMSEA_{0.05}$] = 0.25, lowest overall PC for any fit index). The $RMSEA_{0.05}$ index was also sensitive, although to a much lesser degree, to the factor correlations ($\omega_p^2 = 0.19$) and the skewness of the ordinal variables ($\omega_p^2 = 0.14$), resulting in a noticeable loss in accuracy as the factor correlations and skewness levels increased (e.g., [Table 5.5]: MAE[$RMSEA_{0.05}$] = 0.83, 1.06, and 1.34, for factor correlations of 0.00, 0.30, and 0.50, respectively). The remaining independent variables produced medium, or near medium, main effects ($0.05 \leq \omega_p^2 \leq 0.09$), with the notable characteristic that $RMSEA_{0.05}$ was the only fit index that estimated less factors as the number of variables per factor increased ([Table 5.5]: MBE[$RMSEA_{0.05}$] = -0.68, -0.74, and -0.85, with 4, 8, and 12 variables per factor, respectively).

Regarding the $SRMR_{0.08}$ fit index, the results in Table 5.4 show that it was greatly sensitive to the majority of the independent variables, including the number of factors ($\omega_p^2 = 0.52$), the sample size ($\omega_p^2 = 0.49$), factor loading ($\omega_p^2 = 0.45$), number of variables per factor ($\omega_p^2 = 0.35$), and the skewness of the ordinal variables ($\omega_p^2 = 0.35$). The great sensitivity of this index to many of the independent variables can be illustrated by comparing its performance to that of other fit indices that did not exhibit the same sensitivity. For example, while the $SRMR_{0.08}$ and $TLI_{0.95}$ indices had comparable levels of estimation error with a 1-

factor structure ([Table 5.5]: MAE[$SRMR_{0.08}$] = 0.30 ≈ 0.21 = MAE[$TLI_{0.95}$]), the $SRMR_{0.08}$ index had substantially higher errors of estimation with a 4-factor structure ([Table 5.5]: MAE[$SRMR_{0.08}$] = 1.95 >> 0.89 = MAE [$TLI_{0.95}$]). In addition to its general sensitivity, the $SRMR_{0.08}$ index also displayed an erratic performance across the levels of sample size, exhibiting a strong tendency to *overfactor* with small samples of 100 observations ([Table 5.6]: MBE[$SRMR_{0.08}$] = 1.81), and a strong tendency to *underfactor* with large sample sizes of 1000 ([Table 5.6]: MBE[$SRMR_{0.08}$] = -0.90). This evidently biased performance of $SRMR_{0.08}$ led to accuracy levels that maxed out at medium sample sizes and then decreased as the sample sizes got smaller or larger ([Table 5.6]: PC[$SRMR_{0.08}$] = 0.35, 0.58, and 0.50, with sample sizes of 100, 300, and 1000, respectively).

The performances of the classic factor retention methods, K1, MAP, and PA, give several benchmarks from which to compare the performances of the fit indices. The K1 rule clearly emerged as the worst performing method, showing the greatest sensitivity to the independent variables and producing the lowest levels of accuracy across the majority of the factor levels. As can be seen in Table 5.4, the K1 rule was extremely sensitive to the number of variables per factor (ω_p^2 = 0.95), the factor loading (ω_p^2 = 0.90), number of factors (ω_p^2 = 0.89), sample size (ω_p^2 = 0.83), and, the skewness of the ordinal variables (ω_p^2 = 0.57). In general, K1 was exceedingly inaccurate with large numbers of variables per factor (e.g., [Table 5.5]: with 12 variables per factor, PC[K1] = 0.16 and MAE[K1] = 6.40), low factor loadings, multiple factors, small sample sizes, and highly skewed ordinal variables, while showing a tendency to grossly overestimate the number of factors in these conditions (e.g., [Table 5.6]: with sample sizes of 100, MBE[K1] = 4.90). The MAP method, for its part, exhibited a medium-level performance that included an extreme sensitivity to the factor loading (ω_p^2 = 0.71), number of variables per factor (ω_p^2 = 0.70), and the number of factors (ω_p^2 = 0.53). In particular, MAP performed very poorly with low loadings and/or a few

variables per factor (e.g., [Table 5.5] with 4 variables per factor, PC[MAP] = 0.28 and MAE[MAP] = 1.72), and was excellent with high loadings and many variables per factor (e.g., [Table 5.5] with loadings of 0.70, PC[MAP] = 0.90 and MAE[MAP] = 0.21). In all, MAP and $RMSEA_{0.05}$ were the only methods that consistently underestimated the number of factors.

Table 5.5

Performance of the Factor Retention Methods across the Structure Factors

	Factor Loading			Variables x Factor			Number of Factors			Factor Correlation[a]		
Method	0.40	0.55	0.70	4	8	12	1	2	4	0.00	0.30	0.50
						Proportion Correct						
$CFI_{0.95}$	0.56	0.76	0.86	0.71	0.74	0.73	0.88	0.81	0.60	0.76	0.75	0.59
$TLI_{0.95}$	0.54	0.74	0.86	0.70	0.72	0.72	0.83	0.77	0.62	0.72	0.73	0.63
$RMSEA_{0.02}$	0.47	0.68	0.73	0.61	0.63	0.64	0.68	0.68	0.56	0.67	0.62	0.56
$RMSEA_{0.05}$	0.25	0.53	0.77	0.54	0.52	0.48	0.87	0.66	0.25	0.57	0.46	0.33
$SRMR_{0.08}$	0.27	0.47	0.69	0.52	0.47	0.44	0.80	0.59	0.26	0.52	0.44	0.31
K1	0.08	0.30	0.59	0.54	0.27	0.16	0.58	0.37	0.20	0.28	0.28	0.29
MAP	0.26	0.59	0.90	0.28	0.66	0.81	0.75	0.61	0.49	0.64	0.57	0.45
PA	0.62	0.86	0.95	0.74	0.83	0.86	0.93	0.86	0.72	0.83	0.82	0.71
						Mean Bias Error						
$CFI_{0.95}$	0.32	0.18	0.11	-0.21	0.23	0.59	0.15	0.21	0.22	0.41	0.28	-0.05
$TLI_{0.95}$	0.48	0.34	0.20	-0.07	0.36	0.74	0.21	0.30	0.43	0.53	0.41	0.14
$RMSEA_{0.02}$	-0.12	0.40	0.48	-0.01	0.31	0.47	0.41	0.42	0.04	0.39	0.26	0.04
$RMSEA_{0.05}$	-1.49	-0.74	-0.04	-0.68	-0.74	-0.85	0.16	-0.16	-1.66	-0.67	-0.90	-1.17
$SRMR_{0.08}$	0.06	0.29	0.36	-0.50	0.29	0.92	0.30	0.39	0.06	0.45	0.26	-0.03
K1	5.51	3.44	1.65	0.81	3.39	6.40	0.83	2.29	5.68	4.02	3.99	3.94
MAP	-1.65	-0.87	-0.20	-1.72	-0.67	-0.33	-0.25	-0.57	-1.46	-0.94	-0.97	-1.14
PA	0.04	-0.05	-0.04	-0.23	0.05	0.12	0.07	0.07	-0.13	0.16	0.03	-0.28
						Mean Absolute Error						
$CFI_{0.95}$	0.83	0.43	0.25	0.37	0.43	0.71	0.15	0.30	0.83	0.51	0.50	0.68
$TLI_{0.95}$	0.93	0.51	0.28	0.39	0.51	0.82	0.21	0.37	0.89	0.63	0.59	0.68
$RMSEA_{0.02}$	0.97	0.57	0.49	0.52	0.65	0.86	0.41	0.50	0.94	0.62	0.70	0.83
$RMSEA_{0.05}$	1.57	0.91	0.36	0.83	0.94	1.07	0.16	0.39	1.76	0.83	1.06	1.34
$SRMR_{0.08}$	1.76	1.18	0.64	0.74	1.15	1.69	0.30	0.74	1.95	1.16	1.31	1.55
K1	5.51	3.44	1.65	0.81	3.39	6.40	0.83	2.29	5.68	4.02	3.99	3.94
MAP	1.65	0.87	0.21	1.72	0.68	0.34	0.25	0.58	1.47	0.95	0.97	1.15
PA	0.64	0.21	0.08	0.41	0.28	0.24	0.08	0.19	0.50	0.29	0.29	0.46

Note. A.The results for the 1-factor conditional are not averaged across the factor correlation levels; the numbers next to the fit indexes indicate their cutoff values.

Table 5.6

Performance of the Factor Retention Methods across the Sample Factors

	Sample Size			Skewness			Response Categories			
Method	100	300	1000	0	±1	±2	2	3	4	5
					Proportion Correct					
$CFI_{0.95}$	0.50	0.78	0.90	0.80	0.78	0.60	0.69	0.71	0.74	0.76
$TLI_{0.95}$	0.46	0.76	0.91	0.79	0.77	0.58	0.68	0.70	0.73	0.74
$RMSEA_{0.02}$	0.38	0.66	0.84	0.70	0.67	0.50	0.60	0.62	0.64	0.64
$RMSEA_{0.05}$	0.42	0.53	0.59	0.60	0.55	0.40	0.44	0.49	0.55	0.58
$SRMR_{0.08}$	0.35	0.58	0.50	0.52	0.50	0.40	0.43	0.45	0.51	0.52
K1	0.12	0.29	0.56	0.40	0.35	0.22	0.24	0.31	0.36	0.39
MAP	0.45	0.61	0.68	0.63	0.61	0.50	0.53	0.57	0.60	0.62
PA	0.62	0.84	0.96	0.86	0.83	0.73	0.75	0.80	0.83	0.85
					Mean Bias Error					
$CFI_{0.95}$	0.75	-0.05	-0.09	0.02	0.04	0.55	-0.00	0.26	0.28	0.28
$TLI_{0.95}$	0.99	0.07	-0.03	0.12	0.16	0.74	0.13	0.38	0.42	0.44
$RMSEA_{0.02}$	0.95	-0.03	-0.15	0.09	0.07	0.61	-0.02	0.27	0.35	0.42
$RMSEA_{0.05}$	-0.57	-0.89	-0.82	-0.73	-0.82	-0.72	-1.02	-0.70	-0.68	-0.64
$SRMR_{0.08}$	1.81	-0.20	-0.90	-0.34	-0.06	1.11	0.89	0.23	-0.02	-0.15
K1	4.90	3.68	2.02	2.94	3.30	4.37	4.14	3.64	3.28	3.08
MAP	-1.17	-0.85	-0.71	-0.77	-0.83	-1.12	-1.01	-0.94	-0.86	-0.82
PA	-0.04	0.00	-0.01	-0.05	-0.04	0.04	-0.05	0.03	-0.01	-0.03
					Mean Absolute Error					
$CFI_{0.95}$	1.11	0.28	0.11	0.28	0.34	0.88	0.45	0.57	0.51	0.48
$TLI_{0.95}$	1.30	0.32	0.10	0.32	0.39	1.00	0.50	0.62	0.59	0.58
$RMSEA_{0.02}$	1.30	0.48	0.24	0.43	0.48	1.12	0.62	0.74	0.67	0.67
$RMSEA_{0.05}$	1.09	0.93	0.82	0.79	0.88	1.17	1.08	1.03	0.88	0.80
$SRMR_{0.08}$	1.94	0.74	0.90	0.85	0.99	1.74	1.59	1.17	1.03	0.98
K1	4.90	3.68	2.02	2.94	3.30	4.37	4.14	3.64	3.28	3.08
MAP	1.19	0.85	0.71	0.78	0.83	1.13	1.02	0.94	0.87	0.83
PA	0.64	0.24	0.05	0.21	0.25	0.47	0.40	0.34	0.27	0.23

Note. The numbers next to the fit indices indicate their cutoff values.

The final method, PA, was the least sensitive and most accurate of all the factor retention methods. Only two independent variables had a large effect on the accuracy of PA, the factor loading ($\omega_p^2 = 0.18$) and the sample size ($\omega_p^2 = 0.18$). Also, PA had the highest PC of any method across the different factor levels that were studied (e.g., [Table 5.6]: with 2 response categories, PC[PA] = 0.75, the highest overall PC at this factor level), along with the smallest overall MAE across the various factor levels, with the lone exception coming with 4 variables per factor where $CFI_{0.95}$ had a marginally smaller MAE than PA (e.g., [Table

5.5]: MAE[$CFI_{0.95}$] = 0.37 < 0.41 = MAE[PA]). Moreover, PA was one of only two methods (the other was MAP) that exhibited a coherent behavior across all the factor levels that were manipulated. A factor retention method was considered to exhibit a coherent behavior if it produced more accurate estimates (higher PC and lower MAE) in the more robust conditions, that is, with higher factor loadings, more variables per factor, less factors, smaller factor correlations, larger sample sizes, smaller skewness, and more response categories.

The results for the most salient interactions are presented next in Figures 5.3 to 5.5. The criterion for selecting which interactions to show was that at least the largest effect-sized interaction for each method should be represented in one of the Figures[4]. In the case of the $CFI_{0.95}$, $TLI_{0.95}$, $RMSEA_{0.02}$, and $SRMR_{0.08}$ indices, these were the double interactions of number of variables per factor x sample size ($\omega_p^2 \geq 0.10$) and skewness x sample size ($\omega_p^2 \geq 0.15$). However, because the triple interaction of number of variables per factor x skewness x sample size included both of these double interactions and was also salient amongst these fit indices, it was the one selected for presentation (see Figure 5.3). It should be noted that due to the great similarity in the behavior of the $CFI_{0.95}$, $TLI_{0.95}$, and $RMSEA_{0.02}$ indices, only one was included in Figure 5.3 ($CFI_{0.95}$).

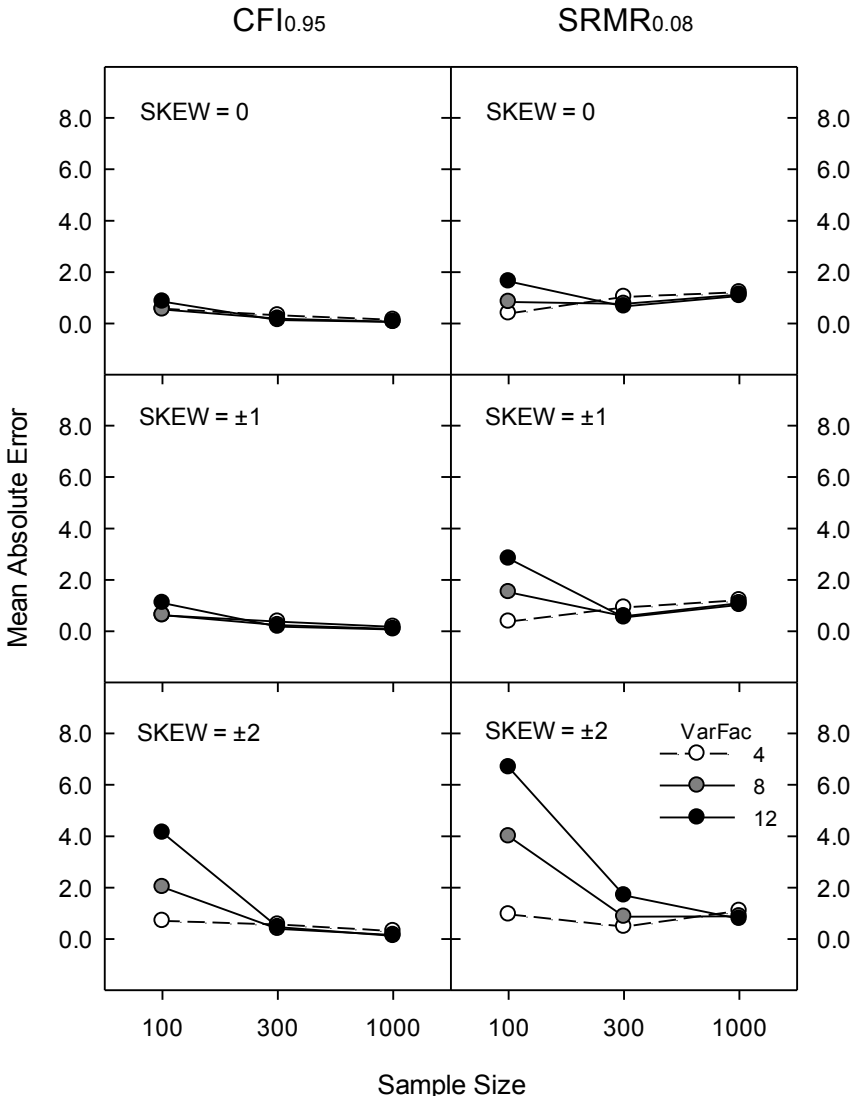
A look at Figure 5.3 reveals that the triple interaction of number of variables per factor x skewness x sample size had a similar impact on the $CFI_{0.95}$ and $SRMR_{0.08}$ indices, although as indicated by their respective effect sizes ($\omega_p^2[CFI_{0.95}] = 0.15$; $\omega_p^2[SRMR_{0.08}] = 0.25$), the pattern was more pronounced for the $SRMR_{0.08}$ index. The triple interaction can be explained as follows: (1) the number of variables per factor x sample size double interaction, which can be observed in most of the 6 blocks of Figure 5.3, arose due to the tendency of the MAE to increase with more variables per factor when the sample size was small (e.g., with sample

[4] This was achieved for all the factor retention methods except K1, where the largest interaction of number of variables per factor x number of factors ($\omega_p^2 = 0.76$) was dropped in favor of the second-largest interaction of factor loading x number of variables per factor ($\omega_p^2 = 0.71$) because of space constraints.

sizes of 100 and skewness of 0, MAE[$SRMR_{0.08}$] = 0.38, 0.84, and 1.64, with 4, 8, and 12 variables per factor, respectively), while subsequently becoming mostly similar for the different number of variables per factor when the sample size was medium or large (e.g., with sample sizes of 1000 and skewness of 0, MAE[$SRMR_{0.08}$] = 1.21, 1.12, and 1.07, with 4, 8, and 12 variables per factor, respectively); and (2) the triple interaction emerged as the tendency of the MAE to increase with more variables per factor with small samples became more pronounced with higher levels of skewness (e.g., with sample sizes of 100 and skewness of ±1, MAE[$CFI_{0.95}$] = 0.62, 0.63, and 1.10, with 4, 8, and 12 variables per factor, respectively, while with sample sizes of 100 and skewness of ±2, MAE[$CFI_{0.95}$] = 0.70, 2.02, and 4.14, with 4, 8, and 12 variables per factor, respectively). Additional results revealed that when the sample size was small, the fit indices affected by this interaction tended to increasingly overfactor with greater numbers of variables per factor (e.g., with sample sizes of 100 and skewness of ±2, MBE[$SRMR_{0.08}$] = 0.83, 3.99, and 6.69, with 4, 8, and 12 variables per factor, respectively).

The double interaction of factor loading x number of factors is shown next in Figure 5.4. This interaction was the most salient for the $RMSEA_{0.05}$ fit index (ω_p^2 = 0.42), the second most salient for the MAP method (ω_p^2 = 0.23), and the fourth most salient for the K1 rule (ω_p^2 = 0.47). As can be seen in Figure 5.4, the factor loading x number of factors interaction emerged for all three methods due to the MAE between the 2 and 4 factor solutions becoming more similar as the factor loadings increased (e.g., MAE[$RMSEA_{0.05}$ with 4 factors] – MAE[$RMSEA_{0.05}$ with 2 factors] = 2.14, 1.60, and 0.40, with factor loadings of 0.40, 0.55 and 0.70, respectively). An inspection of the MBE statistic revealed that the $RMSEA_{0.05}$ index and the MAP method tended to increasingly underfactor with lower factor loadings and more factors (e.g., with 4 factors, MBE[MAP] = -2.58, -1.41, and -0.39, with factor loadings of 0.40, 0.55 and 0.70, respectively). The K1 rule, in contrast,

tended to grossly overestimate the number of factors in these same conditions (e.g., with 4 factors, MBE[K1] = 8.50, 5.66, and 2.88, with factor loadings of 0.40, 0.55 and 0.70, respectively).

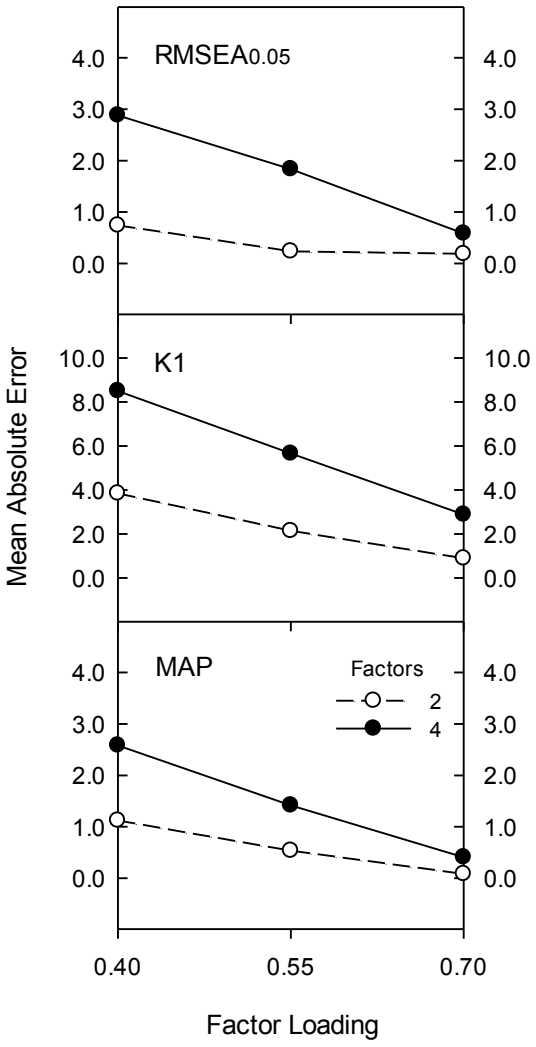


Note. VarFac = variables per factor; the numbers next to the fit indices indicate their cutoff values.

Figure 5.3: *Variables per Factor x Skewness x Sample Size Interaction for CFI and SRMR*

Finally, the factor loading x number of variables per factor interaction is presented in Figure 5.5. This interaction was the most salient for the MAP method ($\omega_p^2 = 0.38$), and the second most salient for the K1 rule ($\omega_p^2 = 0.71$). According to the results shown in Figure 5.5, the factor loading x number of variables per factor interaction had a different effect on

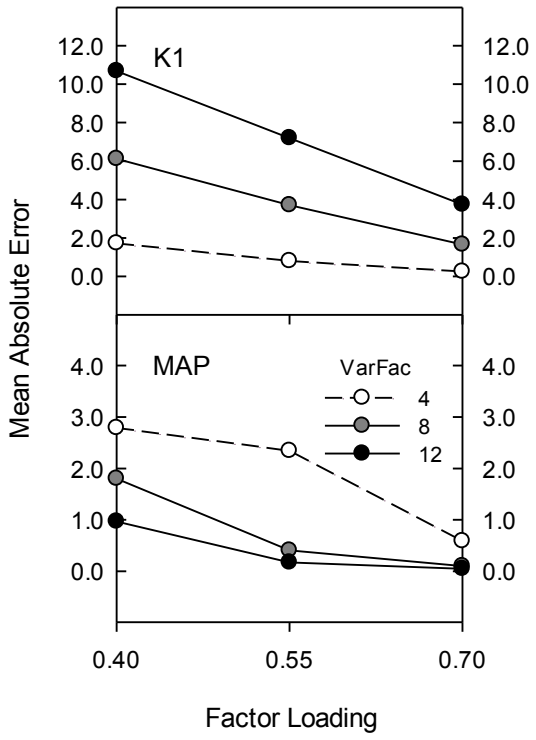
the K1 rule and the MAP method. In the case of the K1 rule, the interaction emerged due to the MAE of the different number of variables per factor becoming more similar as the factor loadings increased (e.g., MAE[K1 with 12 variables per factor] – MAE[K1 with 4 variables per factor] = 8.96, 6.40, and 3.48, with factor loadings of 0.40, 0.55 and 0.70, respectively). It should be noted that the K1 rule always produced smaller errors of estimation with 4 variables per factor, and not with the more robust structures of 8 and 12 variables per factor.



Note. The number next to the fit index indicates its cutoff value.

Figure 5.4: *Factor Loading x Number of Factors Interaction for the RMSEA index and the K1 and MAP Methods*

In terms of the MAP method, on the other hand, the MAE tended to decrease differentially with 4 variables per factor in comparison with 8 or 12 variables per factor, as the factor loadings became higher. While with 4 variables per factor the decrease in MAE was sharpest going from factor loadings of 0.55 to 0.70 (e.g., with 4 variables per factor, MAE[MAP] = 2.79, 2.34, and 0.58, with factor loadings of 0.40, 0.55, and 0.70, respectively), with 8 or 12 variables per factor the decrease was sharpest going from factor loadings of 0.40 to 0.55 (e.g., with 8 variables per factor, MAE[MAP] = 1.80, 0.40, and 0.09, with factor loadings of 0.40, 0.55, and 0.70, respectively). In general, the MAP method produced small errors of estimation with factor loadings of 0.55 as long as the number of variables per factor was medium or large (e.g., with factor loadings of 0.55, MAE[MAP] = 0.40 and 0.16, with 8 and 12 variables per factor, respectively).



Note. VarFac = variables per factor.

Figure 5.5: *Factor Loading x Variables per Factor Interaction for the K1 and MAP Methods*

5.8. Discussion

Researchers in the social sciences have often used fit indices to determine the number of factors as part of a coherent validation strategy in which the fit assessment of the measurement model is not divorced from the dimensionality decision (e.g., Brown et al., 2004; Campbell-Sills et al., 2004; Steiger & Lind, 1980; Tepper & Hoyle, 1996). Recent advances in the areas of exploratory factor analysis (EFA) and structural equation modeling (SEM), such as the development of EFA within a confirmatory factor analysis framework (E/CFA), and of exploratory structural equation modeling (ESEM), have further propelled this synergy between dimensionality and model fit assessment (e.g., Asparouhov & Muthén, 2009; Ferrando & Lorenzo-Seva, 2000; Reichenheim et al., 2011). Researchers are now able to explore unrestricted factorial structures with all the measures of fit and model diagnostics that were only previously available for CFA and SEM models. However, despite this increased use of fit indices to determine the number of factors, the systematic evaluation of their efficacy in this area has so far been scarce (Fabrigar et al., 1999). In order to address this problem, this study used Monte Carlo methods to examine the performance of 4 commonly used fit indices, CFI, TLI, RMSEA, and SRMR, in determining the number of factors for ordinal-level data, typically encountered in the social and behavioral sciences (Flora & Curran, 2004).

An initial issue that had to be addressed was the determination of the optimal cutoff value for each fit index. A review of the CFA and SEM literature revealed that the most recommended cutoff values were 0.95 for the CFI and TLI indices, 0.05 for the RMSEA index, and 0.08 for the SRMR index (Browne & Cudeck, 1992; Chen et al., 2008; Hu & Bentler, 1999; Yu, 2002). However, because these fit indices had not been systematically evaluated in the context of dimensionality assessment, a wide array of cutoff values were studied, which included the ranges of 0.80 to 0.99 with increments of 0.01 for the CFI and

TLI indices, and of 0.20 to 0.01 with decrements of 0.01 for the RMSEA and SRMR indices. The results revealed that the CFI, TLI, and SRMR indices produce an optimal, or near optimal, performance at the recommended cutoff values of 0.95, 0.95, and 0.08, respectively. In contrast, the RMSEA index performs best with a 0.02 cutoff value, rather than the recommended value of 0.05. Based on these results, the fit indices were evaluated at their recommended cutoff values, except for the RMSEA index, which was studied at the optimal and recommended cutoff values of 0.02 ($RMSEA_{0.02}$) and 0.05 ($RMSEA_{0.05}$), respectively.

Once the optimal cutoff values were established for each fit index, the similarity between their dimensionality estimates was examined. According to Pearson's correlation coefficient (r), the estimates of the CFI, TLI, and $RMSEA_{0.02}$ indices have a high level of correspondence ($r \geq 0.87$). This great similarity between the decisions based on the CFI and TLI indices is in line with previous CFA research (e.g., Yu, 2002). On the other hand, the $RMSEA_{0.05}$ index has a low level of correspondence with the other fit indices ($r \leq 0.38$), and only a moderate level of correspondence with $RMSEA_{0.02}$ ($r = 0.52$). The SRMR index, for its part, has a moderate level of correspondence with the CFI, TLI and $RMSEA_{0.02}$ indices ($r \geq 0.68$).

Regarding the overall performance in the determination of the number of factors, the CFI and TLI indices emerged as the two most accurate fit indices, producing correct dimensionality estimates about 70% of the time and a mean absolute deviation from the correct number of factors of approximately half a factor. These indices had also performed well in detecting latent misspecification in CFA models (e.g., Mahler, 2011). Next in line is the $RMSEA_{0.02}$ index, which produces dimensionality estimates that are moderately poorer than those of the CFI and TLI indices, followed by the $RMSEA_{0.05}$ and SRMR indices, which estimated the number of factors correctly only 50% of the time and had a mean absolute error

of approximately one factor. In general, the fit indices tend to overestimate the number of factors, expect for the $RMSEA_{0.05}$ index, which has a strong tendency to underfactor.

In order to evaluate the impact of the independent variables in the performance of the fit indices, separate Analysis of Variance (ANOVA) were performed for each fit index, with the mean absolute error statistic as the dependent variable and the partial omega squared (ω_p^2) statistic as the measure of effect size. According to the ANOVA results, the CFI, TLI, $RMSEA_{0.02}$, and SRMR indices are strongly affected by the sample size ($\omega_p^2 \geq 0.30$), showing a strong tendency to overfactor with small sample sizes (N = 100), as well as by the skewness of the ordinal variables ($\omega_p^2 \geq 0.15$), also overestimating the number of factors with highly skewed ordinal variables (±2). Moreover, these fit indices, and especially the SRMR index, are also strongly affected by the triple interaction of sample size x number of variables per factor x skewness ($\omega_p^2 \geq 0.12$), greatly overestimating the number of factors when small samples (N = 100) are combined with many variables per factor (≥ 8) and highly skewed variables (±2). These results are in line with the findings from the CFA and SEM literature, which shows that the chi-square (χ^2) statistic, as well as the CFI, TLI, RMSEA, and SRMR indices, have a tendency to produce poor fit values and high Type I errors with small samples that contain a large numbers of observed variables and/or variables with high levels of skewness (Hu & Bentler, 1999; Kenny & McCoach, 2003; Lei, 2009; Savalei & Rhemtulla, in press; Sharma et al., 2005; Yu, 2002).

In addition to the previous results, the SRMR index is also highly sensible to the size of the factor loadings and the number of factors ($\omega_p^2 \geq 0.45$), performing very poorly with low factor loadings (0.40) and multiple factors (4). Moreover, this index performs erratically across the levels of sample size, greatly *overfactoring* with small samples (N = 100) and greatly *underfactoring* with large samples (N = 1000), a result that is in line with its performance with CFA models of ordinal variables (e.g., Yu, 2002). The $RMSEA_{0.05}$ index is

also highly sensible to the size of the factor loadings, the number of factors, and the interaction of factor loading x number of factors ($\omega_p^2 \geq 0.42$). In general, the $RMSEA_{0.05}$ index is not able to detect factors with low factor loadings (0.40), resulting in greater errors of estimation as the number of factors with this characteristic increases. This strong underfactoring by the $RMSEA_{0.05}$ index is consistent with the high Type II errors that the index has produced with CFA models that had low factor loadings (e.g., Heene et al., 2011; Mahler, 2011).

As a means of placing the performance of the fit indices into proper context, three classic factor retention methods, the eigenvalue-greater-than-1 rule (K1; Kaiser, 1960), the Minimum Average Partial method (MAP; Velicer, 1976), and Parallel Analysis (PA; Horn, 1965) were also evaluated in the current study. The results indicate that all the fit indices are superior to the K1 rule, which grossly overestimates the number of factors in the majority of the conditions. In addition, the CFI, TLI and $RMSEA_{0.02}$ indices are superior to the MAP method, which is strongly biased to underfactoring with low factor loadings and/or a small number of variables per factor. The $RMSEA_{0.05}$ and SRMR indices, however, are somewhat inferior to the MAP method. Lastly, the PA method is moderately more accurate than the best-performing fit indices, CFI and TLI, consistently producing the highest levels of accuracy across the different research conditions.

Given the combined results of this study, the CFI and TLI indices can be recommended for the dimensionality assessment of ordinal variables, except with very small samples (N = 100). It appears that with ordinal variables, a sample size of 100 is simply not large enough to provide information about the correctness of any particular model when it is combined with many indicators and/or largely skewed variables (Savalei & Rhemtulla, in press). In these conditions, PA is clearly the method of choice, as it is substantially more accurate than any of the other factor retention methods studied here. The $RMSEA_{0.02}$ index, for its part, cannot be

recommended to determine the number of factors due its overall poorer performance in comparison to the CFI and TLI indices, especially with 1- and 2-factor structures, while the $RMSEA_{0.05}$ and SRMR indices should clearly be avoided due to their strong biases and overall erratic performance. A notable finding of this study is that the three factor retention methods that make a comparison against a baseline or null model, CFI, TLI, and PA, performed well, while the four absolute methods, RMSEA, SRMR, K1, and MAP, performed relatively poorly. This result highlights the advantages of incorporating a baseline model in the process of model evaluation and dimensionality assessment, a strategy that explicitly takes into account the effects that the data characteristics (e.g., sample size, number of response categories, skewness levels) might have on the factor structure.

A two-method strategy with the CFI/TLI indices and the PA method appears to be advantageous for the dimensionality assessment of ordinal variables. For example, when the CFI index and the PA method produced the same dimensionality estimate (in 35% of the cases), they had a near-perfect proportion of correct estimates of 0.99, along with a negligible mean absolute error of 0.01. In contrast, when these two methods produced different dimensionality estimates (in 65% of the cases), the proportion of correct estimates was only 0.58 for CFI and 0.71 for PA, while the mean absolute error was 0.77 for CFI and 0.47 for PA. So, a researcher that uses both the CFI index and the PA method can be very confident about the dimensionality estimate when both methods agree, and on the other hand, when they disagree, he/she will be aware that there is a greater possibility of error in the estimation and can take it into account when making the number of factors decision.

There are some limitations in this study that should be noted. First, all the models had perfect simple structure with equal factor loadings, variables per factor, and factor correlations within cases. This strategy is usually preferred for simulation studies because it allows for the generation of data that have perfectly known dimensionalities in the

population. However, these are also idealized models that are not likely to be encountered in practical settings. For this reason, the results from this study should be seen as a best-case scenario. Also, only four fit indices (CFI, TLI, RMSEA, and SRMR) and one estimation method (WLSMV) were evaluated in the current study. Although these fit indices and estimation method are commonly used with EFA and SEM models of ordinal variables, future studies may evaluate the performance of other fit indices and estimation methods, such as the unweighted least squares with mean- and variance-adjusted standard errors estimator (ULSMV; see Savalei & Rhemtulla, in press). Finally, researchers should be aware that the determination of the number of factors is a substantive decision that should not be based solely on statistical information, but also on other criteria such as theory and interpretability of the factor solution (Browne & Cudeck, 1992; Velicer et al., 2000).

6. GENERAL DISCUSSION

The current dissertation used Monte Carlo methods to assess the performance of various factor retention criteria in the dimensionality assessment of ordinal variables. The determination of the number of factors is considered to be a crucial step in exploratory factor analysis (EFA) and structural equation modeling (SEM; Fabrigar, Wegener, MacCallum, & Strahan, 1999; Hayduk & Glaser, 2000; Hayton, Allen, & Scarpello, 2004; Henson & Roberts, 2006; Mulaik & Millsap, 2000; Schmitt, 2011), as errors of under- and overfactoring are likely to result in non-interpretable or unreliable factors (Fava & Velicer, 1992, 1996; Lee & Comrey, 1979; Wood, Tataryn, & Gorsuch, 1996), and can potentially mislead theory development efforts (Fabrigar et al., 1999). In accordance to the importance of the dimensionality decision, considerable research has been undertaken in the past 50 years to develop and evaluate factor retention methods for the assessment of continuous variables (e.g., Cattell, 1966; Horn, 1965; Kaiser, 1960; Velicer, 1976; Velicer, Eaton, & Fava, 2000; Zwick & Velicer, 1986). In contrast, the dimensionality assessment of ordinal variables, which are typically encountered in the social and behavioral sciences (Flora & Curran, 2004), has received much less attention, a situation that could compromise the validity of factor analyses carried out with this type of input data.

Of the many factor retention criteria that have been proposed, a total of seven were studied in this dissertation. Velicer's Minimum Average Partial Method (MAP; Velicer, 1976) and Horn's Parallel Analysis (PA; Horn, 1965) were selected because they had been studied extensively with continuous variables and had emerged as two of the most accurate methods available (Zwick & Velicer, 1982, 1986; Velicer et al., 2000; Hayton et al., 2004; Henson & Roberts, 2006). In addition, the eigenvalue-greater-than-1 criterion or Kaiser's rule (K1; Kaiser, 1960) was also included because it has historically been the most used dimensionality assessment method (Fabrigar et al., 1999; Hayton et al., 2004). These three

factor retention criteria can be considered as "classic" methods, as they were first proposed more than 30 years ago and are based on Principal Component Analysis (PCA) extraction. On the other hand, due to the new developments in the areas of EFA and SEM in the past two decades, researchers have been proposing and using fit indices to determine the number of factors (e.g., Asparouhov & Muthén, 2009; Fabrigar et al., 1999; Ferrando & Lorenzo-Seva, 2000; Floyd & Widaman, 1995; Hoyle & Duvall, 2004; Tepper & Hoyle, 1996). This shift in dimensionality assessment responds to an effort to overcome some of the limitations of the classic factor retention methods, such as the use of PCA extraction or the inability to model correlated errors (Bollen, 2000), and to integrate the number of factors decision within the broader context of SEM modeling (Asparouhov & Muthén, 2009; Ferrando & Lorenzo-Seva, 2000). Following this line, four commonly used fit indices were also selected for evaluation in this study: the Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), Root Mean Square Error of Approximation (RMSEA), and the Standardized Root Mean Square Residual (SRMR). Altogether, these seven methods were chosen to represent a broad range of classic and modern factor retention criteria.

Because the evaluation of the different types of factor retention methods presented unique sets of questions and challenges, three Monte Carlo studies were carried out to adequately assess their performance. The Monte Carlo studies were designed in the following fashion: Study 1 evaluated the performance of the MAP method; Study 2 evaluated the PA method; and Study 3 assessed the performance of the four fit indices and compared it to the performance of the K1 rule, as well as the most accurate MAP and PA variants, which were established in Studies 1 and 2. The following seven "data" factors were manipulated in each study: the factor loading, number of variables per factor, number of factors, factor correlation, sample size, number of response categories, and the level of skewness of the ordinal variables. In addition, other factors such as the extraction method or the type of

correlation matrix were also manipulated depending on the factor retention method that was being studied. A brief summary of the most relevant findings from each study is presented next.

6.1. Most Important Findings from the Monte Carlo Studies

6.1.1. Findings from Study 1: The MAP Method

Study 1 evaluated the performance of the MAP method. Even though this method had been studied extensively with continuous variables (e.g., Peres-Neto, Jackson, & Somers, 2005; Zwick & Velicer, 1982, 1986; Velicer et al., 2000) and had been frequently recommended in best-practice factor analysis guides (Hayton et al., 2004; Henson & Roberts, 2006; Patil, Singh, Mishra, & Donovan, 2008), its effectiveness with ordinal variables had never been examined previously. Aside from the seven data factors that were described previously, two "method" factors were also manipulated in this study: the type of correlation matrix (Pearson or polychoric) and the power that partial correlations were raised to (2^{nd} or 4^{th}). One issue that had to be resolved prior to the study was how to deal with the non-Gramian polychoric matrices, as the MAP method requires Gramian correlation matrices for its computation. Preliminary analyses showed that smoothing the non-Gramian polychoric matrices with the "ridge" procedure (Wothke, 1993) produced good results, and so it was decided to carry out the Monte Carlo study smoothing all the non-Gramian polychoric matrices using this procedure. In order to determine if the emergence and smoothing of non-Gramian polychoric matrices had an impact on the effectiveness of the MAP method, these cases were flagged and analyzed separately.

The results from Study 1 showed that MAP with polychoric correlations is substantially more accurate than MAP with Pearson correlations, regardless of whether the polychoric correlation matrix was originally Gramian or non-Gramian. Although MAP with polychorics

is already moderately superior to MAP with Pearson correlations with unskewed data, the difference becomes substantial as the levels of skewness increase. In terms of the power factor, the results indicated that squaring the partial correlations leads to more accurate estimations than raising them to the 4^{th} power. In general, the most accurate variant of the MAP method is the one with polychoric correlations and the partial correlations squared. On the other hand, all variants of the MAP method exhibit a strong tendency to underestimate the number of factors, especially with low factor loadings and/or a small number of variables per factor, where the MAP method has extremely low levels of accuracy. These latter results are consistent with the findings from previous studies with continuous variables (Zwick & Velicer, 1982, 1986).

A novel finding of Study 1 was the identification of the factor loading x number of variables per factor interaction as having a critical role in the effectiveness of the MAP method. Specifically, with a small number of variables per factor (4), the accuracy of the MAP method does not improve when the factor loadings are raised from a low (0.40) to a medium (0.55) level. In contrast, a similar increase in factor loadings produces a *substantial* improvement in accuracy with medium (8) or large (12) numbers of variables per factor.

6.1.2. Findings from Study 2: Parallel Analysis

Study 2 assessed the performance of the PA method. Parallel Analysis was particular among the factor retention methods studied here in that it was the only method that had received systematic study with ordinal variables. The problem was, however, that the findings of previous studies were inconclusive and often contradictory. Some studies suggested that PA worked better with Pearson correlations (Weng & Cheng, 2005), others that PA performed equally as well with Pearson and polychoric correlations (Cho, Li, & Bandalos, 2009), and still others that PA was ineffective with either type of correlation matrix

(Tran & Formann, 2009). Also, some studies recommended the 95th or 99th percentile of the random eigenvalues (Weng & Cheng, 2005), while others found the mean eigenvalue criteria to be more effective (Cho et al., 2009). In addition, a new version of PA with Minimum Rank Factor Analysis (MRFA) had been recently proposed (Timmerman & Lorenzo-Seva, 2011), and preliminary analyses had shown that it was moderately superior to the original PCA based PA. In order address these inconclusive findings, several characteristics of previous studies were identified that could have led to the discrepant findings, and a large scale Monte Carlo study was designed to assess more comprehensively the performance of PA. The design of Study 2 included the manipulation of three method factors, including the correlation matrix (Pearson or polychoric), the extraction method (PCA or MRFA), and the eigenvalue criteria (the mean of 95th percentile), along with the seven data factors that were manipulated in all the studies.

Similar to the situation in Study 1, a decision had to be made on how to deal with non-Gramian polychoric matrices. Even though PA with PCA extraction could technically be computed with non-Gramian matrices, some authors had resolved to discard these cases (Tran & Formann, 2009), while others had argued that PA could not be interpreted with negative eigenvalues (Timmerman & Lorenzo-Seva, 2011). This issue was tackled in Study 2 with two different strategies: (1) the non-Gramian matrices were smoothed using the Eigenvalue method (Knol & ten Berge, 1989), thus preserving the *strict* rationale of the PA method; and (2) the negative eigenvalues were not given any treatment and theoretical arguments were given on why the *general* rationale of the PA method was still maintained even in these cases. Again, those cases with non-Gramian polychoric matrices were flagged and analyzed separately.

According to the results of Study 2, the emergence of non-Gramian polychoric matrices has no noticeable effects on the accuracy of PA. In addition, smoothing the non-Gramian

matrices and using the original eigenvalues as they are leads the same levels of accuracy, further suggesting that PA is unaffected by the Gramian-status of the polychoric matrices. In terms of the saliency of the method factors, mixed Analysis of Variance (ANOVA) results indicated that the type of correlation matrix and the extraction method has large effects in the accuracy of PA, while the eigenvalue percentile has a medium effect. In general, PA performs best with polychoric correlations, PCA extraction, and the mean eigenvalue criteria. Parallel Analysis with Pearson correlations and PCA extraction performs well with moderately skewed variables (0.0 to ±1.0), but is ineffective with large levels of skewness (±1.5 to ±2.0), due to the emergence of difficulty factors. In addition PA with MRFA has a notably poor performance with a small number of variables per factor (4), and the 95th random eigenvalue percentile has a strong tendency to underfactor with correlated factors. Also, an ANOVA with the best performing PA variant (polychoric correlations + PCA extraction + mean random eigenvalues) revealed that it is mostly affected by the factor loadings and the sample size, both of which have a large impact on the accuracy of this method.

Finally, follow-up analyses showed that the overall best performing PA variant (polychoric correlations + PCA extraction + mean random eigenvalues) is generally the most accurate for all the combinations of the data characteristics that are observable (sample size, number of response categories, and skewness levels), further supporting its use over other PA variants in practice. However, it should be noted that even this best PA variant shows a notable tendency to underestimate the number of factors with small samples, few variables per factor, many factors, and/or high factor correlations.

6.1.3. Findings from Study 3: Fit Indices

Study 3 examined the performance of fit indices. Aside from the research aimed at assessing data unidimensionality (usually done within the context of confirmatory factor analysis [CFA]), no published studies could be identified that evaluated the accuracy of fit indices in determining the number of factors with ordinal variables. Because of this, a largely exploratory approach was taken for study. First, a wide range of cutoff values were evaluated for each index, which included the range of 0.80 to 0.99 in increments of 0.01 for the CFI and TLI indices, as well as the range of 0.20 to 0.01 in decrements of 0.01 for the RMSEA and SRMR indices. Similarly, in order to place the performance of the fit indices into context, three classic factor retention methods were also evaluated in this study: K1, MAP, and PA. Based on the results of previous studies (Zwick & Velicer, 1982, 1986; Velicer et al., 2000), as well as those in Studies 1 and 2 of this dissertation, these three methods were expected to provide low, medium, and high levels of accuracy, respectively, by which to judge and compare the performance of the fit indices. Both the fit indices and the classic methods were computed on the basis of polychoric correlations, and, in the case of the fit indices, a robust weighted least squares estimator appropriate for ordinal data was used (Muthén, du Toit, & Spisic, 1997). Furthermore, the analysis of MAP and PA was limited to best performing variants of each method, which had been determined previously in Studies 1 and 2.

The results of Study 3 revealed that the CFI, TLI, and SRMR indices performed best at the 0.95, 0.95, and 0.08 cutoff values, respectively. These cutoff values correspond to the recommended cutoffs from the CFA and SEM literature (e.g., Hu & Bentler, 1999; Yu, 2002). In contrast, the RMSEA index performed substantially better at the 0.02 cutoff value instead of the usually recommended cutoff of 0.05 (Browne & Cudeck, 1992; Chen, Curran, Bollen, Kirby, & Paxton, 2008; Yu, 2002). Based on these results, the fit indices were further evaluated at their recommended cutoff values, except for the RMSEA index, which was

studied at the optimal and recommended cutoff values of 0.02 ($RMSEA_{0.02}$) and 0.05 ($RMSEA_{0.05}$), respectively.

In terms of the performance of the fit indices in determining the number of factors, the CFI and TLI indices, which produce highly similar dimensionality estimates, are clearly the most accurate fit indices. These indices have also performed well in detecting latent misspecification in CFA models (e.g., Mahler, 2011). Next in line is the $RMSEA_{0.02}$ index, with a performance that is moderately inferior to that of the CFI and TLI indices, followed by the $RMSEA_{0.05}$ and SRMR indices, which perform erratically and have a substantially poorer level of accuracy than that of the CFI and TLI indices. In general, all the fit indices (except $RMSEA_{0.05}$) show a strong tendency to overestimate the number of factors when small samples (N = 100) are combined with many variables per factor (≥ 8) and highly skewed ordinal variables (±2). These results are in line with the findings from the CFA and SEM literature, which shows that the chi-square (χ^2) statistic, as well as the CFI, TLI, RMSEA, and SRMR indices, have a tendency to produce poor fit values and high Type I errors with small samples that contain a large numbers of observed variables and/or variables with high levels of skewness (Hu & Bentler, 1999; Kenny & McCoach, 2003; Lei, 2009; Savalei & Rhemtulla, in press; Sharma, Mukherjee, Kumar, & Dillon, 2005; Yu, 2002). In addition to these results, the $RMSEA_{0.02}$ index has a notably poor performance with 1- and 2-factor structures, due to its tendency to overfactor; the $RMSEA_{0.05}$ index grossly underestimates the number of factors with low factor loadings (0.40), a result that is consistent with the high Type II errors that the index produces with CFA models that have low loadings (e.g., Heene et al., 2011; Mahler, 2011); and the SRMR index is especially erratic across the levels of sample size, greatly *overfactoring* with small samples (N = 100) and greatly *underfactoring* with large samples (N = 1000), a result that is also in line with its performance with CFA models of ordinal variables (e.g., Yu, 2002).

Regarding the placement of the fit indices' performance in comparison to that of the classic factor retention methods, the results indicate that all the fit indices are superior to the K1 rule, which grossly overestimates the number of factors in the majority of the conditions. In addition, the CFI, TLI and $RMSEA_{0.02}$ indices are superior to the MAP method, which is strongly biased to underfactoring with low factor loadings and/or a small number of variables per factor. The $RMSEA_{0.05}$ and SRMR indices, however, are somewhat inferior to the MAP method. Lastly, the PA method is moderately superior to the best-performing fit indices, CFI and TLI, generally producing the highest levels of accuracy across the different research conditions. Also, it should be noted that PA and MAP are the only two methods that exhibit a consistently coherent behavior. A factor retention method is considered to exhibit a coherent behavior if it produces more accurate estimates in the more robust conditions, that is, with higher factor loadings, more variables per factor, less factors, smaller factor correlations, larger sample sizes, smaller skewness, and more response categories.

6.2. Theoretical Context for the Performances of the Retention Methods

The performance of the different factor retention methods can be better understood by putting it into its proper theoretical context. Study 1 showed that the MAP method has a strong bias toward underfactoring with a small number of variables per factor and/or low factor loadings. The reason that the MAP method is biased in these conditions is that it is based on PCA, an extraction method that is known to overestimate the population loadings when the factors are defined by a small number of variables and/or when the population loadings are low (Widaman, 1993). The biases of PCA extraction affect the accuracy of the MAP method in the following manner: the upward bias of the component loadings produces an upward bias in the reproduced (model-implied) correlations, which in turn, results in large partial correlations when components with these characteristics are partialled out from the observed correlation matrix. Then, once these large partial correlations are squared and

averaged they will produce an increase in the MAP criterion, which will result in the factor being discarded and in too few factors being extracted. This situation is further exacerbated when Pearson correlations are used with ordinal data, as the level of association between the variables will be underestimated (Babakus, Ferguson, & Jöreskog, 1987; Bollen & Barb, 1981), resulting in more cases with low communalities. Furthermore, the MAP method also performs better with polychoric correlations because of the emergence of difficulty factors with Pearson correlations (Gorsuch, 1983; Olsson, 1979b). This is evidenced by the overestimation of the number of factors by MAP with Pearson correlations when the ordinal variables have high population loadings and are highly skewed in opposite directions.

Through the results of Study 2 it was established that PA with polychoric correlations is substantially superior to PA with Pearson correlations with largely skewed ordinal variables ($|\text{skewness}| \geq 1.5$). The reason for this difference in accuracy is, like in the case of the MAP method, due to the emergence of difficulty factors with ordinal variables have high population loadings and are highly skewed in opposite directions. However, unlike with the MAP method, PA with Pearson and polychoric correlations perform equally well with moderately skewed variables ($|\text{skewness}| \leq 1.0$). In addition, PA performed very well with PCA extraction and did not exhibit the type of bias of the MAP method with low population loadings and/or a small number of variables per factor.

The main reason for this difference in performance between PA and MAP is that PA is a *comparative* method and MAP is an *absolute* method. For example, when the levels of skewness are not large enough to produce difficulty factors, the reason to still expect biased results with Pearson correlations is the underestimation of the factor loadings. In the case of the MAP method, this is why even with unskewed data the polychoric correlations produce more accurate estimations. In the case of PA, in contrast, because the size of the real and random eigenvalues is affected similarly by this underestimation of the factor loadings with

Pearson correlations, a sort of "cancelling of errors" occurs when these eigenvalues are compared to each other and the bias does not contaminate the dimensionality estimates. This cancelling of errors is also the reason why PA with PCA extraction, unlike MAP, is relatively accurate with low factor loadings and/or a small number of variables per factor. In this case, the overestimation of the loadings that PCA extraction produces works similarly for the real and random data, and thus the *difference* in size between the real and random eigenvalues is mostly free of this bias. The cancelling of errors that occurs with PA, however, does not prevent the emergence of difficulty factors because high factor loadings are necessary for the difficulty factors to emerge (Olsson, 1979b), a condition that can happen with the real data, but which is extremely unlikely to occur with the corresponding random data, due to their population loadings always being zero.

A notable finding of Study 2 is the superior performance of PA with PCA extraction in comparison to PA with MRFA, a common factor extraction method. As was pointed out above, one of the reasons why PA with PCA performs so well is that PA is robust to the biases of PCA extraction due to the comparative nature of the method. Another reason for this superiority is, however, that PA is ill-suited for common factor analysis. There are various problems that arise when PA, originally developed within the PCA framework, is modified for common factor analysis. First, there is the problem of determining which communality estimate to use. Because the communality estimates, and therefore, the size of the eigenvalues, vary as a function of the number of factors extracted, many different solutions can potentially be obtained with a common factor version of PA. The authors that have proposed common factor modifications of PA have resolved this issue by choosing a single communality estimate according to some prespecified criteria. In this line, Humphreys and Ilgen (1969) proposed a PA variant with Principal Axes Factor Analysis (PAFA), where they eigenvalues are computed from a reduced correlation matrix with squared multiple

correlations in the diagonal. Similarly, Timmerman and Lorenzo-Seva (2011) proposed that the communalities for PA with MRFA be computed through the extraction of the maximum number of factors possible (the number of variables [p] minus 1). However, neither the squared multiple correlation (Buja & Eyuboglu, 1992) nor the extraction of $p - 1$ factors are satisfactory solutions for the communality problem, making their use with PA questionable. Second, it is inappropriate to perform factor analysis with random variables, because this type of data violates the assumptions of the common factor model (the variables are uncorrelated to each other at the population level). This situation is likely to produce unacceptable factor solutions, such as those with Heywood cases (Fabrigar et al., 1999), if an iterative estimation method is used. Therefore, although it is desirable to use a common factor extraction method to estimate the number of factors, the PA method is not easily amenable to be modified in this manner.

Regarding the performances of the fit indices, Study 3 showed that all the fit indices except $RMSEA_{0.05}$ have a strong tendency to overfactor when small samples are combined with many variables per factor and/or highly skewed ordinal variables. This behavior is a direct result of the fact that the χ^2 statistic has a tendency to overestimate its theoretical value in this same conditions, leading to poor fit values and high Type I errors (Hu & Bentler, 1999; Kenny & McCoach, 2003; Lei, 2009; Savalei & Rhemtulla, in press; Sharma et al., 2005; Yu, 2002). Also, the poor performance of the RMSEA index can be further understood when put into a broader theoretical context.

A not very well known property of the standard χ^2 based fit indices is that they are more sensitive to misfit when unique variances are small than when they are large (Browne, MacCallum, Kim, Andersen, & Glaser, 2002; Heene, Hilbert, Draxler, Ziegler, & Bühner, 2011). Because of this, indices such as the RMSEA will tend to produce high Type II errors with low factor loadings and high Type I errors with very high factor loadings (Brown et al.,

2002; Heene, et al., 2011; Mahler, 2011). This is the reason why the $RMSEA_{0.05}$ index had such a strong tendency to underfactor with low factor loadings (0.40). This is also part of the reason why $RMSEA_{0.02}$ was also not satisfactory: while the more stringent cutoff value led to substantially better accuracy with low loadings (0.40), it also led to a poorer performance with high loadings (0.70), and its performance would have probably completely fallen apart if even higher loadings had been simulated. In contrast, this problem is less substantial for the incremental fit indices CFI and TLI because with very high factor loadings, for example, the baseline model would have such a poor fit that the overall CFI/TLI values would remain high, signaling a good fit to the data.

6.3. Practical Guidelines

Based on the results of the Monte Carlo studies, the following guidelines are proposed for the dimensionality assessment of ordinal variables. First, PA with polychoric correlations, PCA extraction, and the mean of the random eigenvalues is recommended as the most accurate factor retention method. The following procedure is suggested for the practical implementation of PA: (1) the ordinal variables can be obtained through random column permutations of the real data matrix; and (2) the non-Gramian polychoric matrices may be smoothed using the Eigenvalue method described in Knol & ten Berge (1989) or can be factorized as they are without any transformation.

Second, the CFI and TLI indices with polychoric correlations are also recommended for the determination of the number of factors with ordinal variables, as long as the sample size is not very small (N = 100). Although the CFI index is marginally superior to the TLI index with smaller samples, these two fit indices are mostly interchangeable as they produce very similar dimensionality estimates. The recommended cutoff value for the CFI and TLI indices with ordinal variables is 0.95, while the recommended estimator is weighted least squares

with mean- and variance-adjusted standard errors and χ^2 statistic (WLSMV; Muthén, du Toit, & Spisic, 1997). Because the WLSMV does not require that the sample correlation matrix or the asymptotic covariance matrix be Gramian as part of the estimation procedure, the non-Gramian polychoric matrices can be factorized without any prior smoothing treatment.

Third, a two-method strategy with PA and the CFI/TLI indices appears to be advantageous for the dimensionality assessment of ordinal variables. According to the results of Study 3, when these two types of methods produce the same dimensionality estimate they have near-perfect accuracy, while their level estimation error is considerable when they disagree. So, a researcher that uses both PA and the CFI/TLI indices can be very confident about the dimensionality estimate when both methods agree, and, on the other hand, he/she will be aware that there is a greater possibility of estimation error when they disagree, and can take this information into account when making the number of factors decision.

6.4. Limitations and Future Lines of Study

There are some limitations in this dissertation that should be noted. One of these limitations is that all the models that were simulated had perfect simple structure with equal factor loadings, variables per factor, and factor correlations within cases. This strategy is usually preferred for simulation studies because it allows for the generation of data that have perfectly known dimensionalities in the population, and because it facilitates the interpretation of the effects produced by the independent variables. However, it is also desirable to evaluate the performance of the factor retention methods with simulated data that more closely resembles those found in practice. For example, independent variables like the factor loading or the number of variables per factor could be made to vary within a prespecified range for a particular factor structure. In addition, minor factors can also be simulated, as was done, for example, in Timmerman and Lorenzo-Seva's (2010) evaluation

of the PA method. This latter condition, nevertheless, can be problematic if it is not theoretically clear whether the factor retention methods should detect or not these minor factors. For this reason, a framework must be developed that can delineate unambiguously what should be considered ignorable and non-ignorable variance.

Another limitation of this dissertation is that only four fit indices (CFI, TLI, RMSEA, and SRMR) and one estimation method for these fit indices (WLSMV) were evaluated. Because of this, the findings from this dissertation should be viewed as a preliminary source of information regarding the performance of fit indices in the determination of the number of factors with ordinal variables. Future studies may evaluate the performance of other fit indices and estimation methods, such as the unweighted least squares with mean- and variance-adjusted standard errors estimator (ULSMV), which has been shown to perform well in recent simulation studies (e.g., Savalei & Rhemtulla, in press). In general, the use of fit indices for dimensionality assessment appears to be an open area of research, as more and more researchers are advocating the incorporation of the number of factors decision within the broader context of SEM modeling (Bollen, 2000; Ferrando & Lorenzo-Seva, 2000), and because more appropriate and flexible modeling techniques such as exploratory structural equation modeling (ESEM; Asparouhov & Muthén, 2009) are now available.

An important omission of this dissertation was the exclusion of the scree test (Cattell, 1966), as this method is widely used in practice (Fabrigar et al., 1999), and has performed well enough to be recommended in previous simulation studies (e.g., Zwick & Velicer, 1982, 1986). The scree test was not studied in this dissertation because of the considerable effort involved in procuring multiple experts to evaluate the plot of eigenvalues for large number of data matrices. However, future studies, if possible, should evaluate the accuracy of this method with ordinal variables. Moreover, other factor retention techniques such as the Hull method (Lorenzo-Seva, Timmerman, & Kiers, 2011), which were not available at the

moment the studies in this dissertation were designed, should be further evaluated in the context of ordinal variables. The Hull method is particularly relevant because it is based partly on the information carried by fit indices (Lorenzo-Seva et al., 2011).

Finally, researchers should be aware that the determination of the number of factors is a substantive decision that should not be based solely on statistical information, but also on other criteria such as theory and interpretability of the factor solution (Browne & Cudeck, 1992; Fabrigar et al., 1999; Velicer et al., 2000). In this sense, the estimations of even the best factor retention methods should be regarded as an important component in the dimensionality decision but not as the decision itself.

7. REFERENCES

Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling, 16*, 397-438.

Babakus, E., Ferguson Jr., C. E., Jöreskog, K. G. (1987). The sensitivity of confirmatory maximum likelihood factor analysis to violations of measurement scale and distributional assumptions. *JMR, Journal of Marketing Research, 24*(2), 222-228.

Barrett, P. (2007). Structural equation 148ergen148g: Adjudging model fit. *Personality and Individual Differences, 42*, 815-824.

Bartlett, M. S. (1950). Tests of significance in factor analysis. *British Journal of Psychology*, *3*, 77-85.

Bartlett, M. S. (1951). A further note on tests of significance in factor analysis. *British Journal of Psychology*, *4*(1), 1-2.

Beauducel, A. (2001). Problems with parallel analysis in data sets with oblique simple structure. *MPR online*, *6*(2), 141-157.

Beauducel, A., & Herzberg, P. Y. (2006). On the performance of maximum likelihood versus means and variance adjusted weighted least squares estimation in CFA. *Structural Equation Modeling, 13*, 186-203.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*(2), 238-246.

Bentler, P. M. (1995). *EQS structural equations program manual.* Encino, CA: Multivariate Software.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, *88*, 588–606.

Bollen, K. A. (2000). Modeling strategies: In search of the holy grail. *Structural Equation Modeling, 7*(1), 74-81.

Bollen, K. A. (2002). Latent variables in psychology and the social sciences. *Annual Review of Psychology, 53*(1), 605-634.

Bollen, K. A., & Barb, K. H. (1981). Pearson's r and coarsely categorized measures. *American Sociological Review, 46*(2), 232-239.

Brown, T. A., White, K. S., Forsyth, J. P., & Barlow, D. H. (2004). The structure of perceived emotional control: Psychometric properties of a revised anxiety control questionnaire. *Behavior Therapy*, *35*, 75-99.

Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods and Research, 21*, 230-258.

Browne, M. W., MacCallum, R. C., Kim, C. T., Andersen, B. L., & Glaser, R. (2002). When fit indices and residuals are incompatible. *Psychological Methods, 7,* 403–421.

Buja, A., & Eyuboglu, N. (1992). Remarks on parallel analysis. *Multivariate behavioral research*, *27*(4), 509-540.

Campbell-Sills, L., Liverant, G. I., Brown, T. A. (2004). Psychometric evaluation of the Behavioral Inhibition/Behavioral Activation scales in large sample of outpatients with anxiety and mood disorders. *Psychological Assessment*, *16*(3), 244-254.

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research, 1*(2), 245-276.

Chen, F., Curran, P. J., Bollen, K. A., Kirby, J., & Paxton, P. (2008). An empirical evaluation of the use of fixed cutoff points in RMSEA test statistic in structural equation models. *Sociological Methods and Research, 36*, 462-494.

Cho, S., Li, F., & Bandalos, D. (2009). Accuracy of the parallel analysis procedure with polychoric correlations. *Educational and Psychological Measurement, 69*(5), 748-759.

Cliff, N. (1988). The eigenvalues-greater-than-one rule and the reliability of components. *Psychological Bulletin*, *103*, 276-279.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Comrey, A. L., & Lee, H. B. (1992). *A first course in factor analysis* (2nd ed.). Hillsdale, NJ: Erlbaum.

Crawford, A. V., Green, S. B., Levy, R., Lo, W., Scott, L., Svetina, D., & Thompson, M. S. (2010). Evaluation of parallel analysis methods for determining the number of factors. *Educational and Psychological Measurement*, *70*(6), 885-901.

Dickman, K.W. (1960). Factorial validity of a rating instrument. Unpublished doctoral dissertation, University of Illinois.

Dinno, A. (2009). Implementing Horn's parallel analysis for principal component analysis and factor analysis. *Stata Journal*, *9*(2), 291-298.

Eklöf, H. (2006). Development and validation of scores from an instrument measuring student test-taking motivation. *Educational and Psychological Measurement*, *66*(4), 643-656.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4*(3), 272-299.

Fava, J. L., & Velicer, W. F. (1992). The effects of overextraction on factor and component analysis. *Multivariate Behavioral Research, 27*(3), 387-415.

Fava, J. L., & Velicer, W. F. (1996). The effects of underextraction in factor and component analyses. *Educational and Psychological Measurement, 56*(6), 907-929.

Ferrando, P. J., & Lorenzo-Seva, U. (2000). Unrestricted versus restricted factor analysis of multidimensional test items: Some aspects of the problem and some suggestions. *Psicológica, 21*(3), 301-323.

Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods, 9*(4), 466-491.

Floyd, F J., & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. Psychological Assessment, 7(3), 286-299.

Forero, C., Maydeu-Olivares, A., & Gallardo-Pujol, D. (2009). Factor analysis with ordinal indicators: A Monte Carlo study comparing DWLS and ULS estimation. *Structural Equation Modeling, 16*, 625-641.

Garrido, L E., Abad, F. J., & Ponsoda, V. (2011). Performance of Velicer's minimum average partial factor retention method with categorical variables. *Educational and psychological measurement*, *71*(3), 551-570.

Garrido, L. E., Abad, F. J., & Ponsoda, V. (2012). A new look at Horn's parallel analysis with ordinal variables, *conditionally accepted by Psychological Methods.*

Glorfeld, L W. (1995). An improvement on Horn's parallel analysis methodology for selecting the correct number of factors to retain. *Educational and psychological measurement*, *55*(3), 377-393.

Goodman, L. A., & Kruskal, W. H. (1954). Measures of association for cross classifications. *Journal of the American Statistical Association, 49*, 732-769.

Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale, NJ: Erlbaum.

Guttman, L. (1954). Some necessary conditions for common-factor analysis. *Psychometrika*, *19*(2), 149-161.

Hayduk, L. A., & Glaser, R. (2000). Jiving the four-step, waltzing around factor analysis, and other serious fun. *Structural Equation Modeling, 7,* 1–35.

Hayton, J. C., Allen, D. G., & Scarpello, V. (2004). Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. *Organizational Research Methods, 7*(2), 191-205.

Heene, M., Hilbert, S., Draxler, C., Ziegler, M., & Bühner, M. (2011). Masking misfit in confirmatory factor analysis by increasing unique variances: A cautionary note on the usefulness of cutoff values of fit indices. *Psychological Methods, 16*(3), 319-336.

Henson, R. K., & Roberts, J. K. (2006). Use of exploratory factor analysis in published research: Common errors and some comment on improved practice. *Educational and Psychological Measurement, 66*(3), 393-416.

Herting, J. R., & Costner, H. L. (2000). Another perspective on "the proper number of factors" and the appropriate number of steps. *Structural Equation Modeling, 7*(1), 92-110.

Holgado-Tello, F. P., Chacón-Moscoso, S., Barbero-García, I., & Vila-Abad, E. (2010). Polychoric versus Pearson correlations in exploratory and confirmatory factor analysis of ordinal variables. *Quality and Quantity, 44*(1), 153-166.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30*(2), 179-185.

Hoyle, R. H., & Duvall, J. L. (2004). Determining the number of factors in exploratory and confirmatory factor analysis. In D. Kaplan (Ed.), *Handbook of quantitative methodology for the social sciences* (pp. 301-315). Thousand Oaks, CA: Sage Publications.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1-55.

Humphreys, L. G., & Ilgen, D. R. (1969). Note on a criterion for the number of common factors. *Educational and Psychological Measurement*, *29*, 571–578.

Jöreskog, K G., & Moustaki, I. (2001). Factor analysis of ordinal variables: A comparison of three approaches. *Multivariate Behavioral Research*, *36*(3), 347-387.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement, 20*, 141-151.

Kenny, D. A., & McCoach, D. B. (2003). Effect of the number of variables on measures of fit in structural equation modeling. Structural Equation Modeling, 10, 333–351.

Knol, D. L., & Berger, M. P. (1991). Empirical comparison between factor analysis and multidimensional item response models. *Multivariate Behavioral Research*, *26*(3), 457-477.

Knol, D. L., & ten Berge, J. M. F. (1989). Least-squares approximation of an improper correlation matrix by a proper one. *Psychometrika*, *54*(1), 53-61.

Lai, J.-S., Crane, P. K., & Cella, D. (2006). Factor analysis techniques for assessing sufficient unidimensionality of cancer related fatigue. *Quality of Life Research*, *15*, 1179-1190.

Lee, H. B., & Comrey, A. L. (1979). Distortions in a commonly used factor analytic procedure. *Multivariate Behavioral Research, 14*(3), 301-321.

Lei, P. W. (2009). Evaluating estimation methods for ordinal data in structural equation modeling. *Quality & Quantity, 43*, 495-507.

Lorenzo-Seva, U., & Ferrando, P. J. (2006). FACTOR: A computer program to fit the exploratory factor analysis model. *Behavior Research Methods, Instruments, & Computers*, *38*, 88–91.

Lorenzo-Seva, U., Timmerman, M. E., & Kiers, H. A. L. (2011). The Hull method for selecting the number of common factors. *Multivariate Behavioral Research*, *46*(2), 340-364.

Mahler, C. (2011). The effects of misspecification type and nuisance variables on the behaviors of population fit indices used in structural equation modeling. Unpublished doctoral dissertation, University of British Columbia.

Marsh, H. W., Hau, K. T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling, 11*, 320-341.

Meyers, L.S., Gamst, G., & Guarino, A. (2006). *Applied multivariate research: Design and interpretation.* Thousand Oaks, CA: Sage Publishers.

Mulaik, S. A., & Millsap, R. E. (2000). Doing the four-step right. *Structural Equation Modeling, 7*(1), 36-73.

Murayama, K., Elliot, A. J., & Yamagata, S. (2011). Separation of performance-approach and performance-avoidance achievement goals: A broader analysis. *Journal of Educational Psychology, 103*(1), 238-256.

Muthén, B., & Kaplan, D. (1985). A comparison of some methodologies for the factor analysis of non-normal likert variables. *British Journal of Mathematical and Statistical Psychology, 38*(2), 171-189.

Muthén, B., du Toit, S.H.C. & Spisic, D. (1997). *Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes*. Unpublished technical report.

Nye, C. D., & Drasgow, F. (2011). Assessing goodness of fit: Simple rules of thumb simply do not work. *Organizational Research Methods, 14*(3), 548-570.

O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instruments & Computers, 32*(3), 396-402.

Olsson, U. (1979a). Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika, 44*(4), 443-460.

Olsson, U. (1979b). On the robustness of factor analysis against crude classification of the observations. *Multivariate Behavioral Research, 14*(4), 485-500.

Patil, V., Singh, S., Mishra, S., & Donavan, D. (2008). Efficient theory development and factor retention criteria: Abandon the 'eigenvalue greater than one' criterion. *Journal of Business Research, 61*(2), 162-170.

Peres-Neto, P., Jackson, D., & Somers, K. (2005). How many principal components? Stopping rules for determining the number of non-trivial axes revisited. *Computational Statistics Data Analysis, 49*(4), 974-997.

Preston, C. C., & Colman, A. M. (2000). Optimal number of response categories in rating scales: Reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica, 104*(1), 1-15.

Raykov, T., & Marcoulides, G. A. (2010). *Introduction to psychometric theory*. New York: Routledge.

Reichenheim, M. E., Moraes, C. L., Oliveira, A. S. D., & Lobato, G. (2011). Revisiting the dimensional structure of the Edinburgh Postnatal Depression Scale (EPDS): empirical evidence for a general factor. *BMC Medical Research Methodology*, *11*, 1-25.

Sanne, B., Torsheim, T., Heiervang, E., & Stormark, K. M. (2009). The strengths and difficulties questionnaire in the 156ergen child study: A conceptually and methodically motivated structural analysis. *Psychological Assessment, 21*(3), 352-364.

Savalei, V., & Rhemtulla, M. (in press). The performance of robust test statistics with categorical data. *British Journal of Mathematical and Statistical Psychology*.

Schermelleh-Engel, K., Moosbrugger, H., & Muller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research*, *8*, 23-74.

Schmitt, T. A. (2011). Current methodological considerations in exploratory and confirmatory factor analysis. *Journal of Psychoeducational Assessment, 29*(4), 304-321.

Sharma, S., Mukherjee, S., Kumar, A., & Dillon, W.R. (2005). A simulation study to investigate the use of cutoff values for assessing model fit in covariance structure models. *Journal of Business Research, 58*, 935-943.

Steiger, H. H., & Lind, J. M. (1980, May). *Statistically based tests for the number of common factors*. Paper presented at the annual meeting of the Psychometric Society, Iowa City, IA.

Tepper, K. and Hoyle, R. H. (1996). Latent variable models of need for uniqueness. *Multivariate Behavioral Research*, *31*, 467-493.

Timmerman, M. E., & Lorenzo-Seva, U. (2011). Dimensionality assessment of ordered polytomous items with parallel analysis. *Psychological Methods, 16*(2), 209-220.

Tran, U. S., & Formann, A. K. (2009). Performance of parallel analysis in retrieving unidimensionality in the presence of binary data. *Educational and Psychological Measurement, 69*(1), 50-61.

Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika, 38*(1), 1-10.

Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika, 41*(3), 321-327.

Velicer, W. F., & Jackson, D. N. (1990). Component analysis versus common factor analysis: Some issues in selecting an appropriate procedure. *Multivariate Behavioral Research, 25*(1), 1-28.

Velicer, W. F., Eaton, C. A., & Fava, J. L. (2000). Construct explication through factor or component analysis: A review and evaluation of alternative procedures for determining the number of factors or components. In R. D. Goffin, & E. Helmes (Eds.), *Problems and solutions in human assessment: Honoring Douglas N. Jackson at seventy.* (pp. 41-71). New York, NY, US: Kluwer Academic/Plenum Publishers.

Weng, L. (2004). Impact of the number of response categories and anchor labels on coefficient alpha and test-retest reliability. *Educational and Psychological Measurement, 64*(6), 956-972.

Weng, L., & Cheng, C. (2005). Parallel analysis with unidimensional binary data. *Educational and Psychological Measurement, 65*(5), 697-716.

Widaman, K. F. (1993). Common factor analysis versus principal component analysis: Differential bias in representing model parameters? *Multivariate Behavioral Research, 28*(3), 263-311.

Willmott, C. J., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research*, *30*, 79−82.

Wood, A. M., Maltby, J., Stewart, N., & Joseph, S. (2008). Conceptualizing gratitude and appreciation as a unitary personality trait. *Personality and Individual Differences*, *44*, 619-630.

Wood, J. M., Tataryn, D. J., & Gorsuch, R. L. (1996). Effects of under- and overextraction on principal axis factor analysis with varimax rotation. *Psychological Methods, 1*(4), 354-365.

Wothke, W. (1993). Nonpositive definite matrices in structural modeling. In: *Testing structural equation models*, K. A. Bollen and J. S. Long, eds. Newbury Park, CA: Sage Publications, 256–293.

Yu, C.Y. (2002). *Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes*. Unpublished doctoral dissertation, University of California, Los Angeles.

Zwick, W. R., & Velicer, W. F. (1982). Factors influencing four rules for determining the number of components to retain. *Multivariate Behavioral Research, 17*(2), 253-269.

Zwick, W. R., & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin, 99*(3), 432-442.

8. LIST OF TABLES

9. LIST OF FIGURES

10. APPENDIX 1: GENERAL INTRODUCTION (Spanish Version)

Los investigadores en las ciencias sociales frecuentemente están interesados en el estudio de variables no observables o latentes las cuales son las causas de los comportamientos que ellos observan (Bollen, 2002). Las variables latentes como las aptitudes, emociones, y motivaciones, en el contexto de una teoría bien razonada, tienen el potencial de explicar un amplio rango de procesos comportamentales utilizando un número relativamente pequeño de constructos (Hoyle & Duvall, 2004). Estas variables latentes, o factores, son variables aleatorias cuyas propiedades deben ser inferidas indirectamente utilizando un modelo estadístico que las conecte a variables observables o manifiestas, que se cree son causadas, al menos en parte, por uno o más factores (Mulaik & Millsap, 2000). La herramienta estadística principal para sacar este tipo de inferencias es el análisis factorial, una técnica que tiene como objetivo describir las asociaciones entre un número potencialmente grande de variables observables utilizando un número relativamente pequeño de factores, que son las variables latentes en el modelo estadístico que representan los procesos latentes que han estado operando (Browne & Cudeck, 1992; Hoyle & Duvall, 2004).

Un tópico importante en el análisis factorial exploratorio (EFA, por sus siglas en inglés), y por extensión, en los modelos de ecuaciones estructurales (SEM, por sus siglas en inglés), es la determinación del número de factores a retener para un grupo de variables (Fabrigar, Wegener, MacCallum, & Strahan, 1999; Hayduk & Glaser, 2000; Hayton, Allen, & Scarpello, 2004; Henson & Roberts, 2006; Mulaik & Millsap, 2000; Schmitt, 2011). Extraer muy pocos factores (i.e., subfactorizar) generalmente resulta en un nivel sustancial de error, ya que múltiples factores pueden quedar combinados y las variables pueden saturar en los factores equivocados (Fava & Velicer, 1996; Wood, Tataryn, & Gorsuch, 1996). Alternativamente, extraer demasiados factores (i.e., sobrefactorizar) puede resultar en una "división factorial" así como en factores no interpretables o poco fiables (Fava & Velicer,

1992; Lee & Comrey, 1979; Wood et al., 1996). A pesar de que sobrefactorizar parece tener efectos menos perjudiciales que subfactorizar (Fabrigar et al., 1999), ambos tipos de errores de especificación pueden llevar a una reproducción pobre de la matriz de saturación y a interpretaciones equivocadas (Velicer, Eaton, & Fava, 2000), y pueden potencialmente descarrilar los esfuerzos de desarrollo teórico (Fabrigar et al., 1999; Patil, Singh, Mishra, & Donovan, 2008). Más aún, los investigadores recientemente han enclavado la decisión de dimensionalidad dentro del contexto más amplio de los modelos SEM, ya que una especificación errónea del número de factores puede llevar al rechazo de modelos SEM que por lo demás estarían concebidos correctamente (Mulaik & Millsap, 2000). No obstante, a pesar de que se le ha brindado una atención sustancial a este tópico por más de medio siglo (e.g., Cattell, 1966; Horn, 1965; Kaiser, 1960; Steiger & Lind, 1980; Timmerman & Lorenzo-Seva, 2011; Velicer, 1976; Velicer,et al., 2000), continua siendo una fuente de controversia entre los metodólogos cuantitativos (e.g., Bollen, 2000; Hayduk & Glaser, 2000; Herting & Costner, 2000; Mulaik & Millsap, 2000) y es pobremente comprendido por los investigadores aplicados (Fabrigar et al., 1999; Hayton et al., 2004; Henson & Roberts, 2006).

Existen muchas reglas para determinar el número de factores a retener, y pueden ser divididas en tres categorías (Floyd & Widaman, 1995): pruebas estadísticas, criterios matemáticos y psicométricos, y reglas prácticas.

Pruebas estadísticas. Las pruebas estadísticas para EFA y SEM están disponibles para ciertos métodos de estimación como los de máxima verosimilitud, mínimos cuadrados generalizados y de distribución libre asintótica, y son calculadas como pruebas de significancia de chi-cuadrado (χ^2) sobre la covariación residual entre variables observables luego de haber extraído un determinado número de factores. Si el estadístico chi-cuadrado es significativo el modelo es rechazado a favor de otro modelo con uno o más factores adicionales (Floyd & Widaman, 1995; Fabrigar et al., 1999).

Criterios matemáticos y psicométricos. Estos métodos de retención de factores son usados comúnmente e incluyen los métodos basados en Análisis de Componentes Principales (PCA, por sus siglas en inglés) tales como la regla de los autovalores mayores que 1 o criterio de Kaiser-Guttman (**K1**, Kaiser, 1960), el Análisis Paralelo (**PA**, por sus siglas en inglés; Horn, 1965), y el método de las Mínimas Correlaciones Parciales Medias (**MAP**, por sus siglas en inglés; Velicer, 1976), así como los múltiples índices de ajuste que han sido propuestos en el contexto de los modelos SEM (e.g., Bentler, 1990, 1995; Bentler & Bonnet, 1980; Steiger & Lind, 1980). La regla K1 es el método por defecto en la mayoría de los paquetes estadísticos y está basada en demostraciones a nivel poblacional y en argumentos psicométricos respecto al tamaño de los autovalores para variables no correlacionadas. El Análisis Paralelo, por su parte, puede ser considerado como una "versión muestral" de la regla K1 porque está basado en las mismas demostraciones poblacionales y argumentos psicométricos, pero a diferencia de la regla K1, también toma en cuenta el error muestral, y la "capitalización" por mínimos cuadrados sobre este error en el cálculo de las raíces latentes (Horn, 1965). Adicionalmente, el método MAP emplea una matriz de correlaciones parciales, separando la varianza común de la varianza única y reteniendo los factores que están compuestos principalmente por varianza común (Velicer, 1976; Velicer et al., 2000). Por otro lado, los índices de ajuste, que usualmente están basados en el estadístico χ^2 y miden la "utilidad práctica" de los modelos estadísticos (Bentler & Bonnet, 1980), también han sido recomendados recientemente para la determinación de dimensionalidad (Fabrigar et al., 1999).

Reglas prácticas. Muchos criterios prácticos caen dentro de la rúbrica de reglas prácticas, como la prueba del gráfico de sedimentación (Cattell, 1966), el porcentaje de varianza explicada y el ratio entre el primer y segundo autovalor. De estas reglas de retención de factores, la prueba del gráfico de sedimentación es con diferencia las que más ha sido

utilizada y estudiada (Fabrigar et al., 1999; Hayton et al., 2004), y está basada en la inspección visual del gráfico de los autovalores correspondientes a los factores no rotados.

10.1. Descripción de los Métodos de Retención de Factores más Relevantes

En esta sección se describirán los métodos de retención de factores más relevantes. Estos métodos específicos han sido seleccionados porque han mostrado un buen desempeño en estudios de dimensionalidad tipo Monte Carlo (e.g., Cho, Li, & Bandalos, 2009; Timmerman & Lorenzo-Seva, 2011; Zwick & Velicer, 1982, 1986; Velicer et al., 2000) y/o porque son utilizados frecuentemente en la práctica (Fabrigar et al., 1999; Hayton et al., 2004). Con el fin de ofrecer un contexto histórico, los métodos de retención factorial se presentarán de acuerdo al orden cronológico en que fueron originalmente propuestos.

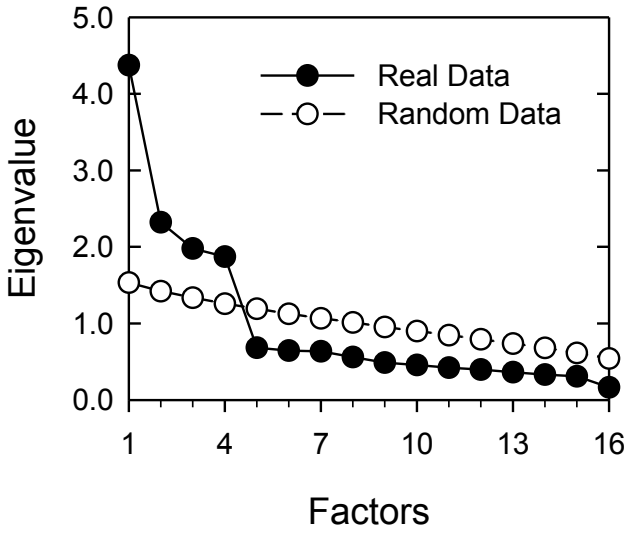
10.1.1. Regla de los autovalores mayores que 1 (1960)

La regla K1 postula que sólo los factores con autovalores mayores que 1 han de ser retenidos. Este criterio se basa en las demostraciones y argumentos de Kaiser (1960) de que el estimado de Guttman (1954) de la raíz latente igual a 1 como el límite inferior para el mínimo rango de una matriz de correlación podría utilizarse como un límite superior psicométrico para el problema del número de factores. La lógica detrás de esta regla es que un factor debe ser capaz de explicar por lo menos tanta varianza como se le otorga a una variable en el espacio estandarizado (Dickman, 1960), y que un umbral de 1 asegura que el componente tendrá un consistencia interna positiva (Kaiser, 1960).

10.1.2. El Análisis Paralelo (1965)

Horn (1965) propuso el método PA sobre la base de las demostraciones y argumentos de Kaiser (1960) y Dickman (1960) de que el estimado de Guttman (1954) de la raíz latente igual a 1 como el límite inferior para el mínimo rango de una matriz de correlación podría utilizarse como un límite superior psicométrico para el problema del número de factores. Dado que las demostraciones para la regla de los autovalores mayores que 1 fueron realizadas sobre estadísticos poblaciones, Horn (1965) argumentó que debido al error muestral, y a la "capitalización" por mínimos cuadrados sobre este error en el cálculo de las raíces latentes, algunos componentes de variables no correlacionadas en la población podrían tener autovalores > 1 al nivel muestral. Por lo tanto, Horn propuso el método PA como una forma de estimar y tomar en cuenta la proporción de varianza que se debía al error muestral y a la capitalización por azar. En este sentido, el método PA puede verse como una alternativa muestral a la regla K1. En vez de retener aquellos factores con autovalores > 1, con el método PA se retienen sólo los factores que tienen autovalores mayores a aquellos generados a partir de variables independientes. El objetivo es contabilizar la capitalización en los autovalores muestrales bajo la hipótesis nula de variables independientes (Buja & Eyuboglu, 1992).

La implementación del método PA involucra la generación de un gran número de matrices de datos aleatorios. Cada matriz es generada con el mismo numero de sujetos y variables que la matriz real bajo evaluación. Entonces, el número de factores se determina comparando los autovalores de la matriz real con la media de los autovalores de las matrices aleatorias (Horn, 1965). Un factor es retenido siempre que su autovalor sea mayor que el autovalor medio del factor aleatorio correspondiente. En la Figura 2.1 se muestra un ejemplo del método PA. En este caso, el método sugiere una estructura de 4 factores.



Nota. Los autovalores para los datos aleatorios representan los valores medios en 100 muestras

Figura 2.1: *Una Ilustración del Análisis Paralelo*

10.1.3. La Prueba del Gráfico de Sedimentación (1966)

La prueba del gráfico de sedimentación (Cattell, 1966) grafica los autovalores de los factores no rotados en un plano de coordenadas y examina la pendiente de la línea que los conecta (Floyd & Widaman, 1995). La lógica de esta prueba es que unos pocos factores principales contienen la mayoría de la varianza, resultando en un "precipicio" en el gráfico de los autovalores, el cual es seguido por una sección baja y plana que describe la varianza relativamente constante y asistemática de los múltiples factores menores (Hayton et al., 2004). El número de factores a retener es igual que al número de factores que no pertenecen a la parte plana del gráfico. En la Figura 2.2 se muestra un ejemplo de la prueba del gráfico de sedimentación. En este caso, la prueba sugiere una estructura de 4 factores, debido a que sólo los primeros 4 factores están localizados antes del "quiebre" en la curva y por lo tanto no pertenecen a la parte plana de la misma.

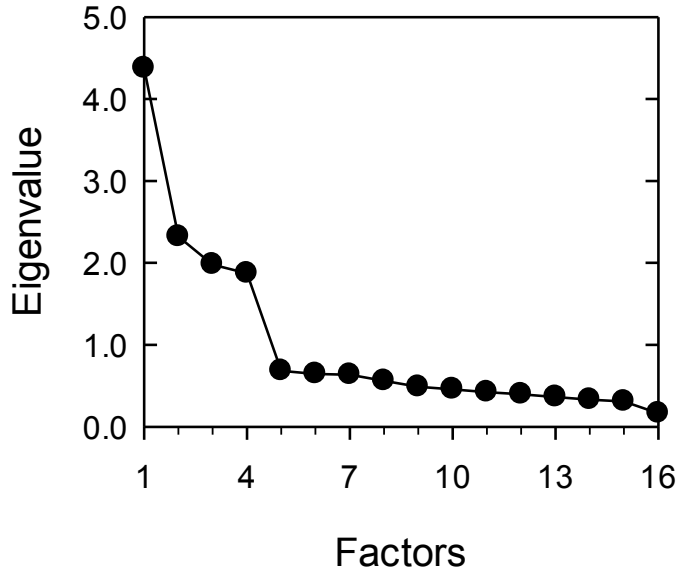


Figura 2.2: *Una Ilustración de la Prueba del Gráfico de Sedimentación*

10.1.4. El Método de las Mínimas Correlaciones Parciales Medias (1976)

El método MAP (Velicer, 1976) fue desarrollado en el contexto del PCA y está basado en la matriz de correlaciones parciales. Cada componente es extraído en secuencia de la matriz de correlaciones y la media de las correlaciones parciales al cuadrado es computada. El número de factores a retener está determinado por el punto en el cual se consigue el valor mínimo de la media de las correlaciones parciales al cuadrado. La lógica de este procedimiento se puede explicar de la siguiente manera: a medida que se extrae varianza común de la matriz de correlaciones de manera sucesiva para cada componente, el criterio MAP irá disminuyendo de valor. En el punto en que la varianza común ha sido removida, el extraer componentes adicionales resultará en la extracción de varianza única y el criterio MAP aumentará de valor. EL procedimiento MAP, por lo tanto, proporciona un punto inequívoco de parada de la extracción de factores al separar la varianza común de la varianza única y al retener sólo aquellos factores que están compuestos principalmente de varianza común. Una ilustración del método MAP se presenta en la Figura 2.3. En este caso, el método

MAP sugiere una estructura de 2 factores ya que el valor mínimo del criterio MAP se obtiene al extraer los primeros dos componentes de la matriz de correlaciones.

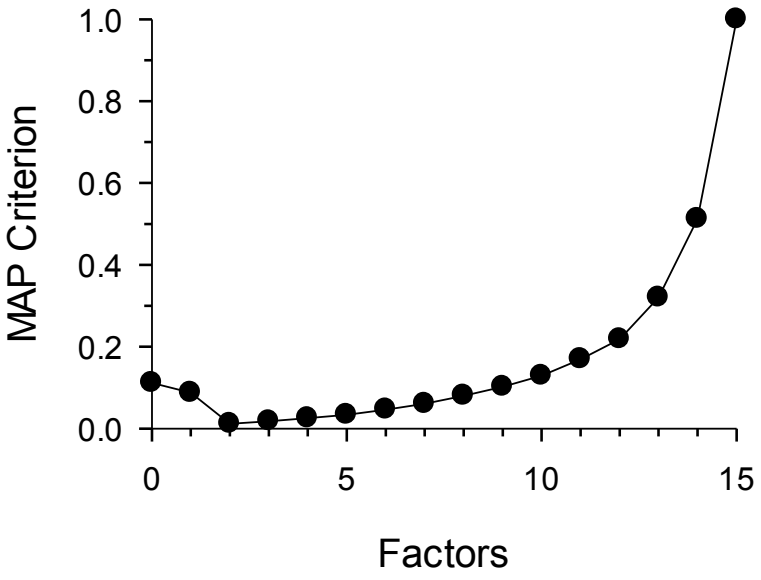


Figura 2.3: Una Ilustración del método MAP.

El procedimiento MAP comienza con el cálculo de la matriz de correlaciones parciales,

$$\mathrm{C}_m = \mathrm{R} - \mathrm{A}_m\mathrm{A}_m^{\mathrm{T}} \tag{10.1}$$

donde C_m es la matriz de covarianzas parciales que resulta de extraer los primeros m componentes de R, R es la matriz de correlación, y A_m es la matriz de saturación para los componentes de 1 a m.

Luego se obtiene la matriz de correlaciones parciales

$$\mathrm{R}_m^* = \mathrm{D}^{-\frac{1}{2}}\,\mathrm{C}_m\,\mathrm{D}^{-\frac{1}{2}} \tag{10.2}$$

donde R_m^* es la matriz de correlaciones parciales, y

$$\mathrm{D} = \mathrm{diag}(\mathrm{C}_m) \tag{10.3}$$

El criterio MAP se obtiene a continuación al promediar los cuadrados de las correlaciones parciales que están contenidas en R_m^*

$$\mathrm{MAP}_m = \sum_{\mathrm{i}=1}^{p} \sum_{\substack{\mathrm{j}=1 \\ \mathrm{i}\neq\mathrm{j}}}^{p} \frac{{\mathrm{r}_{\mathrm{ij}}}_m^{*\,2}}{p(p-1)} \tag{10.4}$$

donde p es el número de variables.

Este procedimiento se repite hasta que $p-1$ componentes se han extraído de R (extraer p componentes resultaría en una matriz de covarianzas parciales nula). Finalmente, Velicer propone una prueba para examinar el primer factor donde se promedian los cuadrados de las correlaciones contenidas en R

$$\mathrm{MAP}_0 = \sum_{\mathrm{i}=1}^{p} \sum_{\substack{\mathrm{j}=1 \\ \mathrm{i}\neq\mathrm{j}}}^{p} \frac{\mathrm{r}_{\mathrm{ij}}{}^2}{p(p-1)} \tag{10.5}$$

si $\mathrm{MAP}_0 < \mathrm{MAP}_1$, ningún factor ha de extraerse.

10.1.5. Índices de Ajuste (1980-)

La forma más común para estimar la dimensionalidad con índices de ajuste es ajustando un modelo de 1 factor a la matriz de correlación y si el índice o índices en cuestión muestran un mal ajuste, más factores son extraídos de forma secuencial hasta que se obtiene un nivel de ajuste aceptable (Asparouhov & Muthén, 2009). El número de factores está determinado por el punto en el cual se obtiene por primera vez un buen nivel de ajuste. A continuación se presenta un breve resumen sobre la lógica de utilizar índices de ajuste para evaluar el ajuste de modelos estadísticos.

La teoría para grandes muestras proporciona una prueba chi-cuadrado (χ^2) de buen ajuste para comparar un modelo contra un modelo alternativo basado en variables correlacionadas (Bentler & Bonett, 1980). El estadístico χ^2 asintótico prueba la hipótesis nula de que la matriz de covarianzas de las variables observadas es igual a la matriz de covarianzas reproducida por el modelo (Hu & Bentler, 1999). Sin embargo, debido a que en las ciencias sociales cualquier modelo es en el mejor de los casos una aproximación a la realidad, se sabe *a priori* que la hipótesis nula de ajuste exacto es falsa, lo que resulta invariablemente, con un tamaño muestral lo suficientemente grande, en el rechace de hasta modelos que se aproximan de manera cercana a la matriz de covarianzas poblacional (Bentler & Bonett, 1980; Browne & Cudeck, 1992; Schermelleh-Engel, Moosbrugger, & Müller, 2003; Yu, 2002). Con el fin de superar estas limitaciones de la prueba estadística χ^2 de ajuste exacto, Bentler & Bonett (1980) propusieron el uso de índices de ajuste *incrementales* como una forma de calcular la cantidad de información ganada al comparar modelos en competencia. Ellos argumentaron que un índice de ganancia de información proporcionaría información importante acerca de la utilidad práctica de modelos en competencia y que debería ser independiente del tamaño muestral y de la información de significación estadística (Bentler & Bonett, 1980). Los índices de ajuste incremental evalúan el grado en el modelo examinado es superior a un modelo alternativo o "base" en la reproducción de la matriz de covarianzas observada. El modelo base es generalmente un modelo nulo en el cual las variables observadas no están correlacionadas (Hu & Bentler, 1999).

Además de los índices de ajuste incremental, también se han desarrollado medidas de ajuste absoluto para evaluar el nivel de adecuación de modelos de variables latentes. Los índices de ajuste *absoluto* miden el grado en que un modelo postulado se corresponde a los datos empíricos. En este caso no se utiliza ningún modelo de referencia para evaluar la ganancia en ajuste, pero se hace una comparación implícita con un modelo saturado que

reproduce exactamente la matriz observada de covarianzas (Hu & Bentler, 1999). Algunos de los índices de ajuste más comúnmente utilizados son el Índice de Ajuste Comparativo (**CFI**, por sus siglas en inglés; Bentler, 1990), el Índice de Tucker-Lewis (**TLI**, por sus siglas en inglés; Tucker & Lewis, 1973), la Raíz Media Cuadrática del Error de Aproximación (**RMSEA**, por sus siglas en inglés; Steiger & Lind, 1980) y la Raíz Media Cuadrática del Residual Estandarizado (**SRMR**, por sus siglas en inglés; Bentler, 1995).

10.2. Revisión de la Literatura de Retención de Factores

10.2.1. Evaluación de la Dimensionalidad de Variables Continuas

Los estudios sobre la precisión de los métodos de retención de factores con variables continuas se han enfocado principalmente en los siguientes 4 métodos "clásicos": la regla K1, la prueba del gráfico de sedimentación, y los métodos PA y MAP. A continuación se presenta un breve resumen de los hallazgos más relevantes de esta literatura.

Zwick & Veicer (1982) utilizaron métodos Monte Carlo para evaluar la efectividad de la regla K1, el test de significancia de Bartlett (Bartlett, 1950, 1951), la prueba del gráfico de sedimentación y el método MAP. Los resultados de su estudio mostraron que la prueba del gráfico de sedimentación era el método más preciso, seguido por el método MAP, que mostró una tendencia a subfactorizar cuando el número de variables por factor era pequeño. A continuación le siguió el test Bartlett, que mostró ser adecuado excepto para los casos con muchas variables, y finalmente estuvo la regla K1, que mostró una tendencia a sobrefactorizar de manera severa el número de factores. En general, el nivel de saturación fue la variable independiente que tuvo el mayor impacto en el desempeño de los métodos de retención de factores evaluados en este estudio.

Utilizando datos más complejos, Zwick y Velicer (1986) encontraron que PA era más efectivo que el método MAP, el cual, a su vez, fue superior a la prueba del gráfico de sedimentación, el test de Bartlett y la regla K1. En este estudio el test de Bartlett resultó muy variable y particularmente sensible al tamaño muestral, el número de variables y el nivel de saturación. En general, este método no pudo ser recomendado debido a su fuerte tendencia a retener factores triviales. Asimismo, y en la línea de los estudios previos, la regla K1 sobrestimó de manera exagerada el número de factores, especialmente con muchas variables observadas. La prueba del gráfico de sedimentación, por su parte, mostró una tendencia hacia la sobrefactorización con niveles bajos de saturación, pero en general, tuvo un desempeño adecuado. A partir de estos resultados, la prueba del gráfico de sedimentación fue recomendada como un método adjunto a los métodos PA y MAP. Respecto al método MAP, en este estudio se observó la misma tendencia hacia la subfactorización con saturaciones bajas y un número pequeño de variables por factor, replicándose así los resultados previos. Por otro lado, PA fue consistentemente el método más preciso, y su desempeño mejoró con aumentos en el tamaño muestral, el nivel de saturación y el número de variables por factor.

Velicer y colegas (2000) condujeron un estudio de simulación donde evaluaron la regla K1, PA y MAP. En este estudio los autores introdujeron una variante en el método MAP al elevar a la cuarta potencia en vez de al cuadrado las correlaciones parciales. Los resultados de su estudio mostraron, nuevamente, que PA era el método más preciso, seguido de cerca por el método MAP. Más aún, la nueva variante del método MAP con la cuarta potencia obtuvo unos resultados que fueron algo más precisos que los obtenidos a partir del método MAP original. Además, y como en estudios previos, la regla K1 sobrestimó de manera exagerada el número factores y fue el método más variable y menos preciso. Finalmente, Peres-Neto, Jackson y Somers (2005) evaluaron el desempeño de 20 reglas de retención de factores en un estudio de simulación y recomendaron, entre otros, a los métodos PA y MAP. En línea con

hallazgos previos, el método MAP mostró una fuerte tendencia hacia la subfactorización en los casos que sugirió un número de factores incorrecto.

Basándose principalmente en estos estudios de simulación con variables continuas, muchos autores han recomendado el uso de los métodos PA y MAP para realizar evaluaciones de dimensionalidad (Fabrigar et al., 1999; Floyd & Widaman, 1995; Hayton et al., 2004; Henson & Roberts, 2006; Patil et al., 2008). Sin embargo, a pesar de que estas recomendaciones están basadas en hallazgos con variables continuas, los investigadores han utilizado frecuentemente estos métodos sin hacer ningún tipo de modificación para la determinación del número de factores con variables ordinales (i.e., Eklöf, 2006; Lai, Crane, & Cella, 2006; Wood, Maltby, & Joseph, 2008), donde su aplicabilidad es cuestionable y su efectividad no ha sido establecida aún.

10.2.2. Evaluación de la Dimensionalidad de Variables Ordinales

Las variables caracterizadas por un nivel de medida ordinal son comunes en muchas investigaciones empíricas dentro de las ciencias sociales y del comportamiento (Flora & Curran, 2004). Una de las situaciones más típicas que generan variables ordinales es el uso de pruebas psicométricas que contienen escalas tipo Likert (e.g., completamente en desacuerdo a completamente de acuerdo). A pesar de que los ítems individuales están diseñados para medir fenómenos continuos, las respuestas observadas son realizaciones discretas con un número pequeño de categorías (Flora & Curran, 2004; Olsson, 1979ª). Si estas variables observadas ordinales son analizadas con métodos estadísticos que asumen distribuciones continuas, existe la posibilidad de que surja un desajuste crítico entre los supuestos que subyacen al modelo estadístico y las características de los datos que se han de analizar (Bollen & Barb, 1981; Flora & Curran, 2004). Debido a esto, usualmente se recomienda que las variables

ordinales sean analizados con métodos diseñados específicamente para este nivel de medida (Savalei & Rhemtulla, en prensa).

Múltiples factores hacen que la evaluación de dimensionalidad con variables ordinales sea más difícil que para variables continuas con distribución normal. Como es bien sabido, la correlación producto-momento de Pearson subestima la fuerza de la relación entre variables ordinales (Babakus, Ferguson, & Jöreskog, 1987; Bollen & Barb, 1981), y puede producir dimensiones artificiales que se conocen como "factores de dificultad" cuando las variables tienen asimetría en direcciones opuestas (Gorsuch, 1983; Olsson, 1979b). Debido a estos sesgos, el coeficiente de correlación policórica ha sido recomendado como una medida de asociación para el análisis factorial de variables ordinales (Flora & Curran, 2004; Jöreskog & Moustaki, 2001). Asumiendo que las variables ordinales son una medida cruda de variables subyacentes con distribución normal bivariante, la correlación policórica es un estimado máximo verosímil de la correlación de Pearson entre las variables subyacentes (Olsson, 1979[a]). Las correlaciones policóricas producen parámetros insesgados tanto para los análisis factoriales exploratorios como confirmatorios (Babakus et al., 1987; Flora & Curran, 2004). A pesar de estas ventajas, sin embargo, las correlaciones policóricas tienen problemas propios también. En particular, frecuentemente producen matrices no-Gramianas (matrices indefinidas con por lo menos un autovalor negativo), tienen errores muestrales grandes y su tiempo de estimación puede ser considerable, propiedades que pueden comprometer su efectividad y aplicabilidad para los métodos de retención de factores como PA o MAP (Timmerman & Lorenzo-Seva, 2011; Tran & Formann, 2009; Weng & Cheng, 2005).

Los estudios de dimensionalidad con variables ordinales se han enfocado fundamentalmente en el Análisis Paralelo. A continuación se presenta un breve resumen de esta literatura. Weng y Cheng (2005) fueron los primeros en estudiar la efectividad de PA con datos binarios unidimensionales al comparar los autovalores empíricos obtenidos a partir de

correlaciones Pearson y tetracóricas con autovalores aleatorios generados a partir de variables con distribución normal multivariante. Los resultados de su estudio de simulación con variables de asimetría positiva mostraron que PA con correlaciones Pearson (PAr) era más efectivo que PA con correlaciones policóricas (Paρ), un hallazgo que atribuyeron a los grandes errores muestrales y el comportamiento inestable de las correlaciones tetracóricas. Un estudio subsecuente por Tran y Formann (2009) extendió la evaluación de PA con datos binarios unidimensionales al simular factores con variables de asimetría positiva y negativa, el caso con mayor probabilidad de producir factores de dificultad para las correlaciones de Pearson. Los resultados de su estudio indicaron que ni PAr ni Paρ podían ser recomendados; en el caso de PAr, debido a un desempeño pobre, y en el caso de Paρ, debido a problemas de aplicabilidad producto de la gran cantidad de matrices policóricas no-Gramianas.

Cho, Li y Bandalos (2009) avanzaron aún más el estudio de PA con variables ordinales al evaluar su desempeño con ítems politómicos de 2 y 3 categorías de respuestas y estructuras multidimensionales de variables con distribución simétrica. En este caso, los autores igualaron el tipo de matriz de correlación utilizada para calcular los autovalores reales y aleatorios y encontraron que PAr era al menos tan efectivo como Paρ. Más recientemente, Timmerman & Lorenzo-Seva (2011) propusieron una nueva versión de PA con Análisis Factorial de Mínimo Rango (MRFA, por sus siglas en inglés), y estudiaron su efectividad y la de la versión original de PA para estructuras multidimensionales de ítems politómicos con asimetría en direcciones opuestas. Según sus resultados, PA con extracción de MRFA fue moderadamente superior a PA con extracción de PCA. Además, estos autores señalaron que Paρ sólo pudo calcularse para el 37% de las matrices de datos debido a problemas de convergencia en el algoritmo utilizado para calcular las correlaciones policóricas o porque el algoritmo de suavizado no produjo una matriz policórica Gramiana. Esta situación llevó a los autores a postular que “los problemas de convergencia del abordaje policórico imposibilitan

su uso generalizado con datos empíricos" y que "pueden ocasionar problemas severos" en la práctica (p. 218).

Como puede observarse a partir de la exposición anterior, el desempeño de PA con variables ordinales ha sido difícil de evaluar debido a que los hallazgos previos han sido inconcluyentes y frecuentemente contradictorios. Por ejemplo, algunos estudios han sugerido que PA funciona mejor con correlaciones Pearson (Weng & Cheng, 2005), otros que PA funciona igual de bien con correlaciones Pearson y policóricas (Cho et al., 2009), y aún otros que PA es inefectivo con cualquier tipo de coeficiente de correlación (Tran & Formann, 2009). Además, diversos estudios han sugerido que PA tiene problemas severos de aplicabilidad con las correlaciones policóricas (Timmerman & Lorenzo-Seva, 2011; Tran & Formann, 2009). Más aún, una nueva versión de PA con extracción de MRFA ha sido propuesta recientemente (Timmerman & Lorenzo-Seva, 2011), y estudios preliminares sugieren que esta versión de PA podría ser moderadamente superior a la versión original con extracción de PCA. Además de estos hallazgos inconsistentes respecto al método PA, otros métodos como el MAP que han sido recomendados frecuentemente (Henson & Roberts, 2006; Velicer, et al., 2000), aún no se han estudiado en el contexto de datos ordinales.

Los índices de ajuste constituyen otro grupo de métodos de retención de factores que podrían utilizarse con variables ordinales. El uso de índices de ajuste para la evaluación de dimensionalidad podría ser ventajoso porque le daría acceso a los investigadores a información diagnóstica importante, como la presencia de errores correlacionados (Bollen, 2000), al momento de tomar la decisión del número de factores. Más aún, podría reducir la necesidad de manipulación *a posteriori* de modelos en las etapas más avanzadas de validación, como la evaluación de modelos SEM completos, debido a una especificación errónea del número de factores (Mulaik & Millsap, 2000). Sin embargo, a pesar de que los índices de ajuste han sido recomendados y utilizados con frecuencia creciente en los últimos

años para la determinación de dimensionalidad (e.g., Asparouhov & Muthén, 2009; Browne & Cudeck, 1992; Hoyle & Duvall, 2004; Fabrigar et al., 1999; Floyd & Widaman, 1995; Steiger & Lind, 1980; Tepper & Hoyle, 1996), su eficacia con variables ordinales aún no ha sido examinada sistemáticamente.

10.3. Objetivos de la Tesis Actual

El objetivo fundamental de la tesis actual es evaluar el desempeño de métodos "clásicos" y "modernos" de retención de factores en la determinación del número de factores para variables ordinales a través de un conjunto integral de factores. Los métodos clásicos a evaluarse incluyen la regla K1, PA y MAP. Por otro lado, los métodos de retención de factores modernos están compuestos por dos índices de ajuste incremental, el CFI y TLI, y de dos índices de ajuste absolutos, RMSEA y SRMR. Un objetivo secundario de la tesis actual es ofrecer guías prácticas que sean claras y fáciles de seguir para los investigadores que trabajan con variables ordinales.

Debido a que muchos de estos métodos de retención de factores requieren modificaciones especiales y presentan retos particulares cuando se usan para determinar la dimensionalidad de datos ordinales, se realizarán tres estudios Monte Carlo para evaluar cada uno de los métodos de manera extensiva. Los objetivos específicos de los tres estudios Monte Carlo se presentan a continuación.

10.3.1. Estudio 1: El Método MAP

El objetivo principal del Estudio 1 es determinar cuál variante del método MAP tiene mejor desempeño con variables ordinales. En esta línea, dos factores de "método" serán manipulados: el tipo de matriz de correlación (Pearson o policórica) y la potencia a la cual son elevadas las correlaciones parciales (la segunda o la cuarta). En este estudio también se

determinará la importancia de un conjunto de variables independientes tales como el nivel de saturación o el número de variables por factor en la efectividad del método MAP. Además, se ofrecerán guías prácticas sobre cómo utilizar el método MAP con variables ordinales. Una cuestión que ha de ser resuelta es qué hacer con las matrices policóricas no-Gramianas, ya que el método MAP no se puede calcular con este tipo de matrices.

10.3.2. Estudio 2: El Análisis Paralelo

El objetivo principal del estudio 2 es determinar cuál variante del método PA tiene un mejor desempeño con variables ordinales. En este caso, tres factores de método serán manipulados: el tipo de matriz de correlación (Pearson o policórica), el método de extracción (PCA o MRFA) y el percentil de los autovalores aleatorios (la media o el percentil 95). Como parte de la evaluación del método PA se hará un intento para identificar y resolver los problemas que han producido los resultados inconsistentes que se encuentran en la literatura. Además, en este estudio también se determinará la importancia de múltiples variables independientes en la efectividad del método PA y se ofrecerán guías prácticas sobre cómo utilizar el método con variables ordinales. Una cuestión que también tendrá que resolverse con PA es qué hacer con las matrices policóricas no-Gramianas, las cuales han sido descartados por muchos autores en el pasado.

10.3.3. Estudio 2: Los Índices de Ajuste

El objetivo fundamental del Estudio 3 es evaluar el desempeño de varios índices de ajuste en la determinación del número de factores con variables ordinales. Con el fin de dar contexto al desempeño de los índices de ajuste, tres métodos clásicos de retención de factores también serán incluidos en la simulación: la regla K1, PA y MAP. En el caso de los métodos PA y MAP, sólo se estudiarán las variantes más efectivas de dichos métodos, las cuales serán establecidas a partir de los Estudios 1 y 2. Además, en este estudio se determinará la

importancia de múltiples variables independientes en la efectividad de los índices de ajuste y se ofrecerán guías prácticas para la determinación de dimensionalidad basadas en métodos de retención de factores clásicos y modernos. Una cuestión que ha de resolverse en este estudio es qué hacer con las soluciones factoriales que no converjan.

11. APPENDIX 2: GENERAL DISCUSSION (Spanish Version)

En la tesis actual se utilizaron métodos Monte Carlo para evaluar el desempeño de diversos criterios de retención de factores para la determinación de dimensionalidad con variables ordinales. La determinación del número de factores es considerada como un decisión crucial en el análisis factorial exploratorio (EFA, por sus siglas en inglés) y los modelos de ecuaciones estructurales (SEM, por sus siglas en inglés; Fabrigar, Wegener, MacCallum, & Strahan, 1999; Hayduk & Glaser, 2000; Hayton, Allen, & Scarpello, 2004; Henson & Roberts, 2006; Mulaik & Millsap, 2000; Schmitt, 2011), ya que errores de sub- y sobrefactorización suelen resultar en factores no-interpretables o de escasa fiabilidad (Fava & Velicer, 1992, 1996; Lee & Comrey, 1979; Wood, Tataryn, & Gorsuch, 1996), y pueden descarrilar los esfuerzos de desarrollo teórico (Fabrigar et al., 1999). En concordancia con la importancia de la decisión de dimensionalidad, una cantidad considerable de investigación se ha llevado a cabo en los últimos 50 años para desarrollar y evaluar métodos de retención de factores para variables continuas (e.g., Cattell, 1966; Horn, 1965; Kaiser, 1960; Velicer, 1976; Velicer, Eaton, & Fava, 2000; Zwick & Velicer, 1986). En contraste, la evaluación de dimensionalidad para variables ordinales, típicamente encontradas en las ciencias sociales y del comportamiento (Flora & Curran, 2004), ha recibido mucha menos atención, una situación que puede comprometer la validez de los análisis factoriales que se llevan a cabo con este tipo de datos.

De los diversos criterios de retención de factores que han sido propuestos, un total de siete fueron estudiados en esta tesis. El Método de la Mínima Correlación Parcial Media de Velicer (MAP, según sus siglas en inglés; Velicer, 1976) y el Análisis Paralelo de Horn (PA, por sus siglas en inglés; Horn, 1965) fueron seleccionados porque habían sido estudiados de manera extensiva con variables continuas y habían emergido como dos de los métodos más precisos (Zwick & Velicer, 1982, 1986; Velicer et al., 2000; Hayton et al., 2004; Henson &

Roberts, 2006). Adicionalmente, el criterio de los autovalores mayores que 1 o regla de Kaiser (K1; Kaiser, 1960) también fue incluido porque históricamente ha sido el método de retención de factores más utilizado (Fabrigar et al., 1999; Hayton et al., 2004). Estos tres métodos de retención de factores pueden ser considerados como métodos "clásicos", ya que fueron propuestos inicialmente hace más de 30 años y están basados en el Análisis de Componentes Principales (PCA, por sus siglas en inglés). Por otro lado, debido a los nuevos desarrollos en las áreas del EFA y los modelos SEM en las últimas dos décadas, los investigadores han propuesto y utilizado los índices de ajuste para determinar el número de factores (e.g. Asparouhov & Muthén, 2009; Fabrigar et al., 1999; Ferrando & Lorenzo-Seva, 2000; Floyd & Widaman, 1995; Hoyle & Duvall, 2004; Tepper & Hoyle, 1996). Este cambio en la evaluación de dimensionalidad responde a un esfuerzo para superar algunas de las limitaciones de los métodos clásicos de retención, como la extracción con PCA o la incapacidad de modelar errores correlacionados (Bollen, 2000), y para integrar la decisión del número de factores dentro del contexto más amplio de los modelos SEM (Asparouhov & Muthén, 2009; Ferrando & Lorenzo-Seva, 2000). Siguiendo esta línea, cuatro índices de ajuste comúnmente utilizados también fueron seleccionados para la evaluación de dimensionalidad: el Índice de Ajuste Comparativo (CFI, por sus siglas en inglés; Bentler, 1990), el Índice de Tucker-Lewis (TLI, por sus siglas en inglés; Tucker & Lewis, 1973), la Raíz Media Cuadrática del Error de Aproximación (RMSEA, por sus siglas en inglés; Steiger & Lind, 1980) y la Raíz Media Cuadrática del Residual Estandarizado (SRMR, por sus siglas en inglés; Bentler, 1995). En conjunto, estos siete métodos fueron seleccionados para representar un amplio rango de criterios clásicos y modernos de retención de factores.

Debido a que la evaluación de los diferentes tipos de métodos de retención de factores presentó preguntas y obstáculos particulares, tres estudios Monte Carlo fueron llevados acabo para evaluar adecuadamente el desempeño de los distintos criterios. Los estudios Monte

Carlo fueron diseñados de la siguiente manera: el Estudio 1 evaluó el desempeño del método MAP; el Estudio 2 evaluó el método PA; y el Estudio 3 evaluó el desempeño de cuatro índices de ajuste y lo comparó con el de la regla K1 y las variantes más efectivas de los métodos PA y MAP, las cuales fueron determinadas previamente a partir de los Estudios 1 y 2. En todos los estudios se manipularon los siguientes siete factores de los datos: el nivel de saturación, el número de variables por factor, el número de factores, la correlación factorial, el tamaño muestral, el número de categorías de respuesta, y el nivel de asimetría de las variables ordinales. Adicionalmente, otros factores como el método de extracción y el tipo de matriz de correlación también fueron manipulados dependiendo del método de retención que se estaba estudiando. Un breve resumen de los hallazgos más relevantes de cada estudio se presenta a continuación.

11.1. Hallazgos más Importantes de los Estudios Monte Carlo

11.1.1. Hallazgos del Estudio 1: El Método MAP

El Estudio 1 evaluó el desempeño del método MAP. A pesar que este método había sido estudiado de manera extensiva con variables continuas (e.g., Peres-Neto, Jackson, & Somers, 2005; Zwick & Velicer, 1982, 1986; Velicer et al., 2000) y que era recomendado frecuentemente en guías de mejores prácticas para el análisis factorial (Hayton et al., 2004; Henson & Roberts, 2006; Patil, Singh, Mishra, & Donovan, 2008), su efectividad con variables ordinales no había sido examinada previamente. Además de los siete factores de los datos que fueron descritos anteriormente, dos factores de "método" también fueron manipulados en este estudio: el tipo de matriz de correlación (Pearson o policórica) y la potencia a la que se elevaron las correlaciones parciales (segunda o cuarta). Una cuestión que tuvo que ser resuelta previo al estudio fue qué hacer con las matrices policóricas no-Gramianas, ya que el método MAP requiere de matrices de correlaciones Gramianas para su

cómputo. En este sentido, análisis preliminares mostraron que suavizar las matrices policóricas no-Gramianas con el procedimiento "ridge" (Wothke, 1993) producía buenos resultados, por lo que se decidió llevar a cabo los estudios Monte Carlo suavizando todas las matrices policóricas no-Gramianas con este procedimiento. Con el fin de determinar si el surgimiento y suavizado de matrices policóricas no-Gramianas tenía algún impacto en la efectividad del método MAP estos casos fueron identificados y analizados por separado.

Los resultados del Estudio 1 mostraron que el método MAP con correlaciones policóricas es sustancialmente más preciso que el MAP con correlaciones Pearson, independientemente de si la matriz policórica es originalmente Gramiana o no-Gramiana. A pesar que MAP con policóricas es de por sí moderadamente superior al MAP con correlaciones Pearson para datos simétricos, la deferencia se convierte en sustancial en la medida en que los niveles de asimetría crecen. Con respecto al factor potencia, los resultados indicaron que elevando las correlaciones parciales al cuadrado se obtienen estimaciones más precisas que elevándolas a la cuarta potencia. En general, la versión más precisa del método MAP es la que utiliza correlaciones policóricas y eleva las correlaciones parciales al cuadrado. Por otro lado, todas las variantes del método MAP exhibieron una fuerte tendencia a subestimar el número de factores, especialmente cuando las saturaciones eran bajas y/o el número de variables por factor era pequeño, condiciones en las cuales el método MAP tiene un nivel de precisión extremadamente bajo. Estos últimos resultados son consistentes con los hallazgos de estudios previos con variables continuas (Zwick & Velicer, 1982, 1986).

Un hallazgo nuevo del Estudio 1 fue la identificación del rol crítico que tiene la interacción de saturaciones x número de variables por factor en la efectividad del método MAP. Específicamente, con un número pequeño de variables por factor (4), la precisión del método MAP no mejora cuando las saturaciones aumentan de un nivel bajo (0.40) a un nivel

medio (0.55). En cambio, un aumento similar en las saturaciones produce un incremento sustancial en precisión con un nivel medio (8) o grande (12) de variables por factor.

11.1.2. Hallazgos del Estudio 2: El Análisis Paralelo

El Estudio 2 evaluó el desempeño del método PA. El Análisis Paralelo tenía la particularidad entre los métodos de retención de haber sido evaluado previamente de manera sistemática con variables ordinales. El problema era, sin embargo, que los hallazgos de los estudios previos eran inconcluyentes y frecuentemente contradictorios. Algunos estudios sugerían que PA funcionaba mejor con correlaciones Pearson (Weng & Cheng, 2005), otros que PA funcionaba igual de bien con correlaciones Pearson y policóricas (Cho, Li, & Bandalos, 2009), y aún otros que PA era inefectivo con cualquier tipo de matriz de correlación (Tran & Formann, 2009). Además, algunos estudios recomendaban el percentil 95 o 99 de los autovalores aleatorios (Weng & Cheng, 2005), mientras otros encontraron que la media era más efectiva (Cho et al., 2009). También, una versión nueva de PA con extracción de Análisis Factorial de Rango Mínimo (MRFA, por sus siglas en inglés) había sido propuesta recientemente (Timmerman & Lorenzo-Seva, 2011), y análisis preliminares habían mostrado que era moderadamente superior a la versión original de PA con PCA. Con el objetivo de abordar estos hallazgos inconcluyentes, diversas características de los estudios previos fueron identificadas que podrían haber producido los resultados contradictorios, y se diseñó un estudio Monte Carlo a gran escala para evaluar de manera extensiva el desempeño de PA. En el diseño del Estudio 2 se incorporaron tres factores relacionados al método PA, entre los que estaban la matriz de correlación (Pearson o policórica), el método de extracción (PCA o MRFA), y el percentil de los autovalores aleatorios (la media o el percentil 95), además de los siete factores de los datos que se manipularon en todos los estudios.

De manera similar al Estudio 1, fue necesario tomar una decisión sobre las matrices policóricas no-Gramianas. A pesar de que PA con extracción PCA podía ser calculado con matrices no-Gramianas, algunos autores habían decidido eliminar estos casos (Trann & Formann, 2009), mientras que otros habían argumentado que PA no podía ser interpretado con autovalores negativos (Timmerman & Lorenzo-Seva, 2011). Esta problemática fue abordada en el Estudio 2 a partir de dos estrategias distintas: (1) las matrices no-Gramianas fueron suavizadas utilizando el método de los Autovalores (Knol & ten Berge, 1989), así preservando la lógica *estricta* del método PA; y (2) no se les dio ningún tratamiento a los autovalores negativos al tiempo que se ofrecieron argumentos teóricos de por qué la lógica *general* del método PA se mantenía aún en estos casos. Nuevamente, los casos con matrices policóricas no-Gramianas fueron identificados y analizados por separado.

Según los resultados del Estudio 2, las matrices no-Gramianas no producen efectos notables en la efectividad de PA. Adicionalmente, suavizar las matrices no-Gramianas y utilizar los autovalores originales sin tratamiento producen los mismos niveles de precisión, sugiriendo de nuevo que PA no se ve afectado por el estatus-Gramiano de la matriz policórica. En cuanto al impacto de los factores relacionados al método PA, Análisis de Varianza (ANOVA) indicaron que el tipo de matriz de correlación y el método de extracción tienen un efecto grande en la precisión de PA, mientras que el percentil de los autovalores aleatorios tiene un efecto medio. En general, PA funciona mejor con correlaciones policóricas, extracción PCA y la media de los autovalores aleatorios. El análisis paralelo con correlaciones Pearson y extracción PCA funciona bien cuando los variables tienen un nivel de asimetría moderado (0.0 a ±1.0), pero es inefectivo con niveles grandes de asimetría (±1.5 a ±2.0), debido al surgimiento de factores de dificultad. Adicionalmente, PA con extracción MRFA muestra un desempeño pobre con un número pequeño de variables por factor (4) y el percentil 95 de los autovalores aleatorios tiene una fuerte tendencia a subfactorizar cuando

los factores están correlacionados. Además, un ANOVA con la mejor variante de PA (correlaciones policóricas + extracción PCA + media de los autovalores aleatorios) mostró que se veía afectado principalmente por las saturaciones y el tamaño muestral, las cuales alcanzaron un nivel de impacto grande en la efectividad del método PA.

Finalmente, análisis adicionales mostraron que la variante más efectiva de PA a nivel general (correlaciones policóricas + extracción PCA + media de los autovalores aleatorios) es generalmente la más efectiva para todas las combinaciones de las características de los datos que son observables (tamaño muestral, número de categorías de respuesta, y nivel de asimetría), apoyando aún más su uso en la práctica sobre otras variantes. Sin embargo, se debe destacar que inclusive esta variante de PA muestra una tendencia notable a subestimar el número de factores con tamaños muestrales pequeños, pocas variables por factor, muchos factores, y/o correlaciones factoriales altas.

11.2. Guías Prácticas

Utilizando como base los estudios Monte Carlo de esta tesis se proponen las siguientes guías para la evaluación de dimensionalidad con variables ordinales. Primero, PA con correlaciones policóricas, extracción PCA y la media de los autovalores aleatorios se recomienda como el método de retención de factores más preciso. Se sugiere el siguiente procedimiento para la implementación práctica de PA: (1) las variables ordinales pueden obtenerse a partir de permutaciones aleatorias de las columnas de la matriz de datos reales; y (2) las matrices policóricas no-Gramianas pueden ser suavizadas utilizando el método del Autovalor descrito por Knol y ten Berge (1989) o pueden ser factorizadas como están sin darles ningún tratamiento.

Segundo, los índices CFI y TLI con correlaciones policóricas también se recomiendan para la determinación del número de factores con variables ordinales, siempre que el tamaño

muestral no sea demasiado pequeño (N = 100). A pesar de que el índice CFI es marginalmente superior al índice TLI con muestras pequeñas, estos dos índices de ajuste son mayormente intercambiables ya que producen estimaciones de dimensionalidad muy similares. El punto de corte recomendado para los índices CFI y TLI con variables ordinales es 0.95, mientras que el estimador recomendado es el de mínimos cuadrados ponderados con media- y varianza-ajustada de los errores típicos y el estadístico chi-cuadrado (WLSMV, por sus siglas en inglés; Muthén, du Toit, & Spisic, 1997). Debido a que el WLSMV no requiere que la matriz de correlación muestral o la matriz de covarianzas asintótica sean Gramianas como parte del proceso de estimación, las matrices policóricas no-Gramianas pueden ser factorizadas sin necesidad de darles ningún tratamiento previo.

Tercero, parece que una estrategia combinada de dos tipos de métodos con PA y los índices CFI/TLI puede ser ventajosa para la evaluación de dimensionalidad con variables ordinales. Según los resultados del Estudio 3, cuando estos dos tipos de métodos producen el mismo estimado de dimensionalidad ellos tienen un nivel de acierto casi perfecto, mientras que su nivel de error es considerable cuando sus estimados no son iguales. Por lo tanto, un investigador que utiliza ambos PA y los índices CFI/TLI puede tener mucha confianza en el estimado de dimensionalidad cuando ambos métodos están de acuerdo, y, en cambio, él/ella estará consciente de que hay una mayor posibilidad de error de estimación cuando los métodos están en desacuerdo, y puede tomar esta información en consideración cuando tome la decisión acerca del número de factores.